Transforming Unstructured Data in IT Project: A Comparative Study of Zero-Shot and Generative AI Text Classification

Cai Tung-Iersloy[1*] Worapat Paireekreng[2] Nantika Prinyapol[3]

[1*,2,3]*College of Innovative Technology and Engineering, Dhurakij Pundit University, Bangkok, Thailand*

*Corresponding Author. E-mail address: 637191110006@dpu.ac.th

Received: 7 August 2024; Revised: 7 October 2024; Accepted: 8 October 2024
Published online: 26 December 2025

Abstract

In today's world, we have a lot of messy, unorganized data from things like comments, interviews, and images. This is especially true in IT projects, where there's often too much information to handle easily. Our study looks at how we can turn this messy data into useful numbers and insights using smart computer programs. We tested two main methods: Zero-Shot Text Classification and Generative AI Text Classification. Zero-Shot is like having a smart assistant that can sort information without needing examples first. Generative AI is more like having a creative writer who can come up with new examples to help sort information. We asked 42 participants with experience in working with unstructured data to answer some questions, then used these methods to analyze their answers. We found that Zero-Shot works better for information that has clear patterns, while Generative AI is good at handling more complex or unclear information. Our results show that choosing the right method can make a big difference in how well we understand and use the data. Zero-Shot was about 15% more accurate for well-organized information, while Generative AI was 20% better at dealing with complex, messy data. This research helps companies and researchers choose the best way to make sense of their data, especially in IT projects where there's often too much information to handle manually.

Keywords: Qualitative data, Unstructured data, Zero-shot text classification

I. INTRODUCTION

The proliferation of unstructured qualitative data in the digital era poses significant challenges for effective analysis, particularly in information technology (IT) projects. With approximately 80% of data remaining unstructured [1], this issue extends across various sectors. IT projects, encompassing software development, network upgrades, and cybersecurity implementations, are at the forefront of managing this data deluge.

Transforming unstructured data into quantitative insights involves unitization, categorization, and coding. Large Language Models (LLMs) like GPT-3 have shown promise in automating this process [2], though careful application is crucial to avoid inaccuracies. Zero-Shot Classification offers powerful tools for categorizing data without task-specific training. The CLORE (Classification by LOgical REasoning) framework [3] and semantic knowledge integration techniques [4] exemplify this approach, leveraging logical reasoning on natural language explanations for effective classification.

This paper explores the application of LLMs and Zero-Shot Classification in revolutionizing data management for unstructured data in IT projects. By leveraging AI to transform qualitative data into quantitative insights, organizations can enhance decision-making and data analysis efficiency [5], unlocking the economic and innovative potential of unstructured data.

II. LITERATURE REVIEW

This review examines Zero-Shot Text Classification and Generative AI Text Classification as key methodologies for transforming unstructured data in IT projects. IT projects, in this context, refer to technology-driven initiatives within organizations involving the development, implementation, or maintenance of information systems and digital infrastructure. These encompass activities such as software development, network upgrades, data management systems, cybersecurity implementations, cloud migrations, ERP and CRM system deployments, and mobile application development. Such projects often generate and handle vast amounts of unstructured data, making them ideal candidates for advanced AI-driven analysis techniques. Zero-Shot Classification utilizes existing knowledge to categorize text without task-specific training [3], [4], while Generative AI Text Classification employs large language models to generate and classify text, adapting to complex patterns [2]. Recent advancements by Zhang et al. [4] and Abburi et al. [2] have enhanced these techniques, with Ye et al. [6] expanding Zero-Shot capabilities using pre-trained models and prompt learning. Despite the potential demonstrated by Yin et al. [7] and Brown et al. [8] in NLP tasks and human-like text generation, the comparative effectiveness of these techniques in transforming qualitative IT project data into quantitative insights remains unexplored. This study aims to bridge this research gap, potentially revolutionizing unstructured data processing and analysis in IT project management.

A. Transforming Qualitative Data into Quantitative Results

The process of converting qualitative data into quantitative insights is crucial for effective analysis in various fields. Srnka and Koeszegi [9] propose a systematic approach involving structured data collection, rigorous transcription, and well-defined categorization and coding processes. This method emphasizes the scientific measurement of qualitative data, although it acknowledges that natural human understanding and open-ended responses often yield powerful qualitative insights. Modern AI techniques, including Zero-Shot Text Classification [3], Large Language Models (LLMs) [2], and Generative AI Text Classification [4], offer promising solutions to automate and enhance this transformation process, addressing the challenges posed by large datasets and the need for efficient data analysis.

As Table 1 illustrates, there are various mixed research designs for integrating qualitative and quantitative approaches:

Table 1: Qualitative-Quantitative Research Designs: Types, Descriptions, and Aims

Research	Description
Sequential two-studies design	**Description:** Qualitative data and quantitative data are collected and analyzed in sequential order. **Aim:** Investigate under-researched fields, develop hypotheses or create instruments for subsequent quantitative measurement, or provide explanations.
Concurrent two-studies design	**Description:** Both quantitative and qualitative data are collected and analyzed in separate procedures. **Aim:**Cross-validate or corroborate findings of the two approaches.
Integrated elaboration design	**Description:** Quantitative data is analyzed using qualitative procedures. **Aim:** Investigate and understand the problem in depth, derive new theoretical insights.
Integrated generalization design	**Description:** Qualitative material is collected and transformed into categorical data for further quantitative analysis. **Aim:** Derive both theory and generalizable results.

B. Zero-Shot Text Classification

Zero-Shot Text Classification is an advanced technique for categorizing text without task-specific training. The CLORE (Classification by LOgical REasoning) framework, introduced by Han et al. [3], exemplifies this approach through two main stages:

1) Logical Parsing: Breaking down explanations into logical structures to identify relevant attributes.

2) Logical Reasoning: Matching these attributes to input data for classification scoring.

This method demonstrates superior performance in tasks requiring high-level logical reasoning and offers improved interpretability compared to baseline models. It also shows robustness against linguistic biases, making it versatile for various classification [3]. Zhang et al. [4] further enhanced this approach by proposing a two-phase framework that integrates semantic knowledge:

1) Coarse-Grained Classification: Using a traditional classifier to determine if an input belongs to seen or unseen classes.

2) Fine-Grained Classification: Employing a zero-shot classifier with semantic knowledge-based feature augmentation for more precise categorization.

This framework significantly improves zero-shot text classification, particularly for domains with evolving or diverse classification needs. However, it may face challenges with tasks requiring new data generation or handling highly complex patterns, highlighting the potential complementary nature of zero-shot and generative approaches in addressing diverse text classification challenges.

C. Generative AI Text Classification

Generative AI text classification differs from zero-shot text classification by focusing on the generation and classification of new text data, rather than relying solely on existing data and logical reasoning. Generative AI models, such as those combining multiple LLMs, can create new content and classify it based on learned patterns. [2] explore the application of ensemble models combining multiple Large Language Models (LLMs) for generative AI text classification.

Table 2 from shows the results of various models for the Binary-English task. The ensemble with voting classifier outperformed individual models, demonstrating the strength of combining outputs from multiple LLMs.

Table 2: Results for the Binary-English task

Model	Accuracy	Macro F1	Precision	Recall
deberta-large				
	0.62	0.546	0.783	0.61
xlm-r-100langs-bert-base-nli-stsb-mean-tokens				
	0.647	0.592	0.782	0.639
roberta-base-openai-detector				
	0.679	0.636	0.805	0.671
xlm-roberta-large-xnli-anli				
	0.618	0.543	0.782	0.608
roberta-large				
	0.623	0.551	0.784	0.613
Ensemble with Voting classifier				
	0.751	0.733	0.826	0.745

Generative AI text classification excels in handling complex and nuanced tasks by leveraging LLMs' capabilities to create new data, filling gaps in existing datasets. However, this approach requires careful management to avoid generating inaccurate or misleading information. Recent research by Abi Akl [10] demonstrates the synergy between traditional ML techniques and LLM-generated data, achieving a Macro-F1 score of 88.401% on seed data with a 1.5% performance increase when augmented by LLM-generated data. This combination not only enhances text classification accuracy but also provides a robust pathway for extracting and utilizing unstructured data within larger frameworks. The effectiveness of LLMs in creating robust text embeddings further expands their application, such as in code comment classification. Models like ChatGPT showcase the potential of generating additional data to improve classical machine learning systems, highlighting the versatility of AI in handling diverse unstructured data types.

D. The Framework of Extracting Unstructured Usage for Big Data Platform

To test AI techniques in transforming qualitative data, [11] propose a framework for extracting and utilizing unstructured data in organizations, distinguishing between structured and unstructured data. Chasupa and Paireekreng's framework [11] enhances decision-making by converting qualitative inquiries into quantitative measurements through a 4x4 questionnaire, ensuring effective data transformation.

Table 3: Unstructured Big Data Extracting Model

		Unstructured Data Form			
		Object	Event	Command	Outcome
Unstructured Data Activity	Property				
	Cause				
	Truth				
	Journey				

Literature Review Summary: The discussed research highlights that for transforming qualitative data into quantitative results, the framework provided by Srnka and is effective for smaller datasets [9]. However, when dealing with large datasets, retrospective interviews, or comments, significant challenges and limitations arise, especially in research grounded in qualitative data or in extracting data from new domains. Replacing the three stages of Srnka and Koeszegi's framework with AI techniques can enhance the diversity and scope of analysis and research.

Zero-Shot Text Classification and Generative AI Text Classification offer promising solutions. Zhang et al. show that Zero-Shot Text Classification is highly effective for tasks requiring logical parsing and semantic knowledge integration. This method is suitable for structured, interconnected data. On the other hand, Generative AI Text Classification, as explored by Abburi et al. [2] and Abi Akl [10], is adept at handling complex and nuanced tasks by generating new data and enhancing traditional ML systems with LLM-generated data. This method is beneficial for data requiring interpretation or filling in gaps.

To determine the most appropriate method and validate the research hypothesis, both tools will be employed to compare their results. The goal is to ascertain which method is more accurate and suitable for converting large-scale qualitative data into quantitative insights and to identify the specific contexts in which each method excels. This approach will provide a comprehensive understanding of the effectiveness of AI in managing and transforming qualitative data.

III. RESEARCH METHODOLOGY

The proposed methodology aims to transform qualitative data into quantitative insights using two AI techniques: Zero-Shot Text Classification and Generative AI Text Classification. The methodology will involve the following key steps as Figure 1.

1) Data Collection: Gather qualitative data through detailed questionnaires and narrative responses.

2) Data Wrangling: Clean and transform the collected data to ensure consistency and usability.

3) Data Preprocessing: Prepare the qualitative data for analysis by categorizing and coding it.

4) Chose AI Technique is Model Application.

5) Zero-Shot Text Classification: Apply Zero-Shot Text Classification to categorize the qualitative data without task-specific training.

6) Generative AI Text Classification: Apply Generative AI Text Classification using ensemble models combining multiple LLMs to generate new data and classify it.

7) Comparison and Analysis: Compare the results of both AI techniques to determine their accuracy, effectiveness, and suitability for various data contexts.

8) Evaluation: Evaluate the performance of each AI technique and identify the specific contexts in which each method excels.

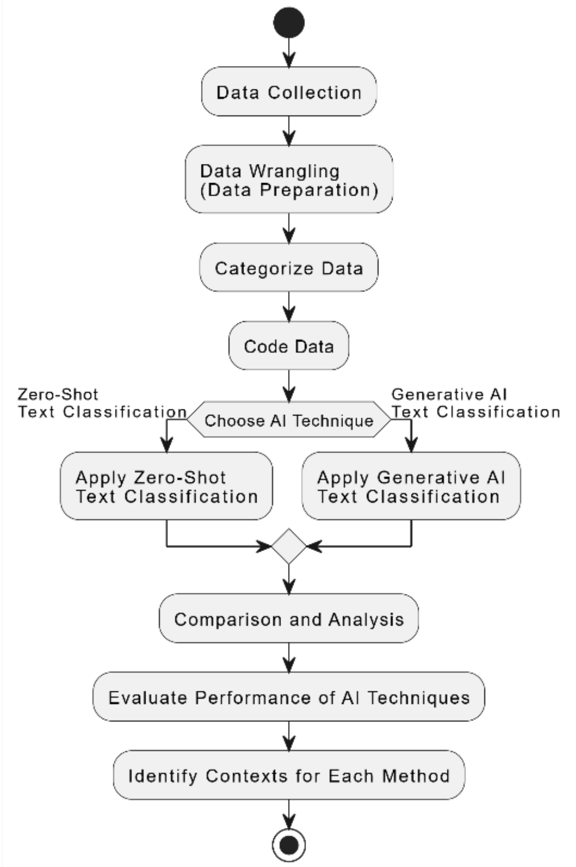

Figure 1: Research Methodology

A. Data Collection

The data collection process will involve gathering qualitative data from participants involved in projects related to unstructured data. The participants will be selected from nine different types of projects and six IT job positions from Talance [12]. The total population size will be 54 individuals, and the sample size will be calculated using Cochran's formula to ensure representativeness.

B. Data Wrangling.

Data wrangling is vital for preparing qualitative data for analysis, involving tasks like handling missing values, correcting inconsistencies, and segmenting narrative responses. Language models, particularly large ones, assist in these tasks through few-shot or zero-shot inference [13] Participants are drawn from projects in various domains, including 1) OCR projects, 2) NLP projects, 3) Paperless document management systems, 4) Sentiment and opinion analysis, 5) Image and video analysis, 6) Audio data analysis, 7) Social media data analysis, and 8) Recommendation systems. The appropriate sample size, determined using Cochran's formula by bin Ahmad and binti Halim [14]. with an 85% confidence level, is 42 participants as shown in Equation 1.

$$n = \frac{207.36}{1+\left(\frac{207.36-1}{54}\right)} = \frac{207.36}{1+3.83} = \frac{207.36}{4.83} \approx 42.95 \qquad (1)$$

IV. RESULTS

This section presents the findings of the study, focusing on the transformation of qualitative data into quantitative insights using Zero-Shot Text Classification and Generative AI Text Classification. The objectives of the research are reiterated to provide context for the results. A detailed analysis of each table is presented, followed by an integrated discussion of all results to provide a comprehensive overview of the study's findings.

A. Data Preparation and Initial Observations.

To ensure clarity and diversity in the questions, the 4×4 model from "The Framework of Extracting Unstructured Usage for Big Data Platform" was extended to include dimensions of project management. This addition enriched the data with project management perspectives, resulting in a comprehensive 4×4×4 framework. The added dimensions are.

1) Time: Represents different phases of a project, helping to identify when events or activities occur and estimate the duration required for each phase.

2) Priority: Indicates the importance level of data or activities, aiding in prioritizing urgency and resource allocation efficiently.

3) Resources: Represents the resources required for operations or analysis, such as data, technology, personnel, or capital.

4) Stakeholder: Identifies groups affected by or influential to the project, helping to manage expectations and requirements clearly [1].

This created a set of questions with enhanced depth, forming a 4×4×4 framework as illustrated in Figure 1.

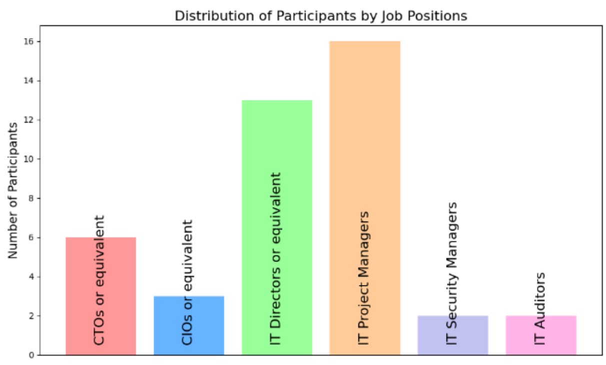

Figure 2: Distribution of Participants by Job Positions

The study involved participants from nine types of unstructured data projects and six IT job positions, totaling 54 individuals. Using Cochran's formula, the sample size was determined to be 42-43 participants, resulting in a final selection of 42 participants. The distribution of participants is as follows: 6 CTOs or equivalent, 3 CIOs or equivalent, 13 IT Directors or equivalent, 16 IT Project Managers, 2 IT Security Managers, and 2 IT Auditors, as shown in Figure 2.

Participant demographics included 31 Southeast Asians, 3 Indians, 4 Chinese, 1 Japanese, 2 Thai working in the U.S., and 1 Russian, highlighting a diverse range of perspectives. All were experienced in IT systems related to unstructured data. The data collection process involved:

1) Initial Contact via various channels.

2) Interviews using a 16-question framework from "The Framework of Extracting Unstructured Usage for Big Data Platform,"

3) Data Processing with a 6-point scoring system,

4) Human Review of qualitative responses, and

5) Data Verification with participants. This process is depicted in Figure 3.

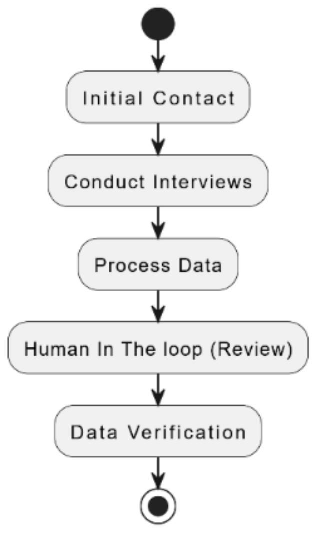

Figure 3: Data collection process

B. Data Separation with RAG Technique.

The collected data from 42 participants, comprising responses to 16 questions, covered various aspects including physical characteristics, Time, Priority, Resources, and Stakeholder dimensions. To process this data, interviews were transcribed and refined using GPT, while written responses were converted to text. The resulting text was then segmented and mapped to relevant concepts from "The Framework of Extracting Unstructured Usage for Big Data Platform" using Retrieval Augmented Generation (RAG) with GPT-4. This approach minimized hallucinations and enhanced accuracy in the data processing [15]. The context-tuned planner, based on the work of Anantha et al. (2024), achieved a notable AST-based Plan Accuracy of 85.24%, while significantly reducing hallucinations to 0.93%. This improvement underscores the importance of context integration in enhancing the accuracy and reliability of the planning process for unstructured data analysis. The Context-tuned Upper Bound was selected for performing RAG, ensuring that sentence segmentation remained within the contextual framework. This choice was crucial for preventing excessive hallucinations and ensuring more accurate and contextually relevant segmentation, which is essential for analyzing unstructured data in IT projects [15]. As Table 4 illustrates:

Table 4: End-to-end Planner Evaluation

Setting	AST-based Plan Acc	Exact Match	Hallucination
Lower Bound	43.77	39.45	2.59
RAG-based Planner	76.39	58.12	1.76
Context-tuned RAG Planner	85.24	67.33	0.93
Upper Bound	91.47	72.65	0.85
Context-tuned Upper Bound	91.62	72.84	0.53

C. Application of AI Techniques Zero-shot Text Classification

The Zero-Shot Text Classification approach uses Python tools to interpret context, with interviews conducted mainly in Thai and English. The code was adapted for both languages, using facebook/bart-large-mnli for English and joeddav/xlm-roberta-large-xnli for Thai. The joeddav/xlm-roberta-large-xnli model, a multilingual extension of RoBERTa, supports Thai and is trained for cross-lingual inference (XNLI). In contrast, facebook/bart-large-mnli is a monolingual model designed for English, utilizing the BART architecture for text generation and transformation.

Trained for MNLI (Multi-Genre Natural Language Inference), which involves understanding linguistic inference in various English genres. Reasons for Choosing joeddav/xlm-roberta-large-xnli for Thai:

The joeddav/xlm-roberta-large-xnli model, trained for Thai, outperforms the facebook/bart-large-mnli model, which only supports English, by directly processing Thai text and reducing translation errors. For example, in response to "What documents need to be scanned and how is the workflow prioritized, including resources and time management?" the model effectively applies Zero-Shot Text Classification, crossing Object x Property and incorporating Project Management dimensions like Time, Priority, Resources, and Stakeholder." The Thai response is: "ใบแจ้งหนี้ที่เราต้องการแปลงเป็นรูปแบบดิจิทัล ซึ่งเป็นสิ่งสำคัญสำหรับระบบการจัดการเอกสารของเรา จะได้รับทุกวันก่อน 12:00 ผ่านไปรษณีย์ไทย ซึ่งจะมาปนกับจดหมายอื่นๆ เราจำเป็นต้องคัดแยกเพื่อนำไปสแกนเอกสารและ OCR เพื่อให้สามารถประมวลผลและจัดการข้อมูลได้อย่างมีประสิทธิภาพภายใน 2 ชั่วโมง"

This translates to English: "The invoices we need to digitize, which are crucial for our document management system, are received daily before 12:00 PM via Thai Post, mixed with other mail. We need to sort them for scanning and optical character recognition to efficiently process and manage the data within 2 hours."

When the example context is run using the Zero-Shot model joeddav/xlm-roberta-large-xnli, the results are as shown in Figure 4. For comparison, when the same context translated into English is run using the facebook/bart-large-mnli model, the results differ as shown in Table 5.

```
from transformers import pipeline

# Load the Zero-Shot Classification Pipeline with a multilingual model
classifier = pipeline("zero-shot-classification", model="joeddav/xlm-roberta-large-xnli")

# The sentence to classify
sentence = "ใบแจ้งหนี้ที่เราต้องการแปลงเป็นรูปแบบดิจิทัล ซึ่งเป็นสิ่งสำคัญสำหรับระบบการจัดการเอกสารของเรา จะไ

# Labels to check for relevance
labels = ["Object", "Property", "Time", "Priority", "Resources", "Stakeholder"]

# Perform classification for each label and check results
threshold = 0.10
for label in labels:
    result = classifier(sentence, [label])
    if result['scores'][0] >= threshold:
        print(f"Relevant to {label}")
    else:
        print(f"Not relevant to {label}")

Some weights of the model checkpoint at joeddav/xlm-roberta-large-xnli were not used when
- This IS expected if you are initializing XLMRobertaForSequenceClassification from the ch
- This IS NOT expected if you are initializing XLMRobertaForSequenceClassification from th
Not relevant to Object
Relevant to Property
Relevant to Time
Relevant to Priority
Relevant to Resources
Not relevant to Stakeholder
```

Figure 4: Zero-Shot model joeddav/xlm-roberta-large-xnli

Table 5: Compare for Thai response and English

	joeddav/xlm-roberta-large-xnli	facebook/bart-large-mnli
Object	Not relevant	Relevant
Property	Relevant	Relevant
Time	Relevant	Relevant
Priority	Relevant	Relevant
Resources	Relevant	Relevant
Stakeholder	Not relevant	Relevant

This table compares the performance of joeddav/xlm-roberta-large-xnli and facebook/bart-large-mnli models in classifying Thai and English text respectively. The results show that the Thai model (joeddav/xlm-roberta-large-xnli) performs better in identifying 'Property', 'Time', and 'Resources' categories, while the English model (facebook/bart-large-mnli) excels in 'Object' and 'Stakeholder' categories. This difference in performance highlights the importance of language-specific models in multilingual text classification tasks.

D. Generative AI Text Classification

To ensure accurate classification by ChatGPT, the PARTS Framework (Persona, Action, Result, Target, and Style) was employed for prompting, providing a structured approach to generating precise and context-appropriate prompts for AI models. This framework defines the AI's role, specifies the task, outlines expected outcomes, identifies the intended audience, and determines the response tone, leading to improved classification results. A portion of the prompt used is shown in Figure 5, with the results summarized in Table 6.

```
**Purpose:**
Read and understand the text, then determine if it
relates to the given set of keywords.

**Audience:**
Computer program processing relevance as Y or N for
further analysis.

**Requirements:**
Accepts text in both Thai and English, written formally.

**Tone:**
Polite

**Scope:**
1. First, ask for the "text" that needs to be analyzed
for relationships.
2. Then, ask for the set of keywords for classification,
explaining that they will be used to check relevance.
3. Once both the text and the set of keywords are
obtained, understand their context.
4. If they are related, explain how they are related.
```

Figure 5: Sample Prompt PARTS Framework

Table 6: Results of Generative AI Text Classification

Keyword	Relevance
Object	Relevant
Property	Not relevant
Time	Relevant
Priority	Relevant
Resources	Relevant
Stakeholder	Not relevant

The results demonstrate the Generative AI model's effectiveness in identifying certain categories ('Object', 'Time', 'Priority', and 'Resources') while showing limitations in others ('Property' and 'Stakeholder'). This pattern suggests that the Generative AI approach may be more suitable for tasks focusing on temporal and resource-related aspects of data in IT projects.

E. Comparative Analysis and Validation and Human-in-the-Loop.

When both techniques, Zero-Shot and Generative AI, were used and validated by the respondents, the results were obtained as shown in Table 7. Reduce the size of the table, "Relevant" is denoted as "Y" and "Not relevant" as "N".

Table 7: Comparative Results

	joeddav	facebook	Generative AI	Human In The Loop
Object	Y	Y	Y	Y
Property	Y	Y	N	Y
Time	Y	Y	Y	Y
Priority	N	Y	Y	N
Resources	Y	Y	Y	Y
Stakeholder	Y	Y	N	Y

This comparative analysis provides insights into the strengths and limitations of each AI approach relative to human judgment. The results suggest that while AI models show promising performance, there are still areas where they differ from human classification, particularly in categories like 'Priority' and 'Stakeholder'.

F. Summary of Key Findings

This analysis used Human In The Loop data to benchmark the reliability of three models: joeddav/xlm-roberta-large-xnli, facebook/bart-large-mnli, and a Generative model. A Proportion Test assessed each model's reliability against a standard. The proportion of matches was calculated, followed by a weighted average. Z-scores were determined and compared to a Z-critical value of 1.96 for a 95% confidence level. With a sample size of 42 participants answering 16 questions (672 responses), results are summarized in Table 7.

G. Summary of Comparative Analysis and Validation

1) joeddav/xlm-roberta-large-xnli shows high reliability, as its Z-value is 3.41, exceeding the Z-critical value (1.96).

2) Generative also shows high reliability, as its Z-value is -2.68, which is significant at a 95% confidence level (in the negative direction indicating significant deviation).

3) facebook/bart-large-mnli is not statistically significant, with a Z-value of -0.73, which is below the Z-critical value (1.96).

Table 8: Z-Test Hypothesis Results

Model	Matches	Mismatches	Matches (%)	Mismatches (%)	Z-value	Significant
joeddav	624	48	93%	7%	3.41	Yes
facebook	590	82	88%	12%	-0.73	No
Generativ	574	98	85%	15%	-2.68	Yes

The analysis shows that the joeddav/xlm-roberta-large-xnli and Generative models are more reliable than the Human In The Loop standard, while the facebook/bart-large-mnli model falls short in comparison.

These results provide quantitative evidence for the relative strengths of each approach. The joeddav model shows the highest reliability, suggesting its potential for multilingual applications in IT projects. The Generative model's significant result, albeit in the negative direction, indicates its unique approach to classification tasks.

H. Integrated Analysis of Results

The collective analysis of Tables 4-8 provides a comprehensive view of the performance and reliability of different AI techniques in transforming unstructured qualitative data into quantitative insights within IT projects. The RAG technique (Table 4) demonstrates the importance of context in improving accuracy and reducing errors. The comparison of language-specific models (Table 5) highlights the nuanced differences in processing Thai and English text, which is crucial for multilingual IT environments. The Generative AI results (Table 6) show its strength in certain categories, particularly those related to time and resources, which are often critical in IT project management. The human-in-the-loop validation (Table 7) offers insights into how well these AI models align with human judgment, an important factor in practical applications. Finally, the statistical analysis (Table 8) provides a quantitative basis for comparing the reliability of these different approaches.

Overall, these results suggest that while each method has its strengths, the joeddav/xlm-roberta-large-xnli model demonstrates the highest overall reliability for this specific task in IT projects. However, the strong performance of the Generative AI model in certain categories indicates its potential for specialized applications within IT project management, particularly in areas focusing on temporal and resource-related data. These findings have significant implications for choosing appropriate AI techniques for different types of unstructured data in various IT project contexts.

V. DISCUSSION

The findings from this study highlight the efficacy and reliability of different AI models in transforming qualitative data into quantitative insights. The Zero-Shot Text Classification and Generative AI Text Classification techniques were rigorously evaluated using a sample of 42 participants who provided 672 responses. These responses were analyzed and validated against Human In The Loop data to benchmark the performance of three AI models: joeddav/xlm-roberta-large-xnli, facebook/bart-large-mnli, and Generative.

A. Zero-Shot Text Classification

The Zero-Shot Text Classification approach, especially with the joeddav/xlm-roberta-large-xnli model, showed high reliability. With a Z-value of 3.41, well above the critical 1.96, this model effectively handled Thai text, which was the majority of the data. Its multilingual capabilities make it a robust tool for organizations dealing with multilingual unstructured data, offering reliable classifications across different languages and contexts.

B. Generative AI Text Classification.

The Generative AI Text Classification technique demonstrated strong performance, with a Z-value of -2.68. Using the PARTS Framework for prompting, it generated accurate classifications validated by human respondents. This model's ability to contextualize data with minimal pre-training indicates its potential for handling complex datasets. Its performance suggests that generative approaches offer flexible, adaptive solutions for real-time data classification, suitable for dynamic environments.

C. Comparison & Implications

Comparing the AI techniques reveals key strengths: joeddav/xlm-roberta-large-xnli excels in structured, multilingual tasks with a high reliability, while the Generative AI model is more adaptable for dynamic environments. The facebook/bart-large-mnli model, with a Z-value of -0.73, lacks the robustness of the other models, suggesting limited utility and potential areas for future enhancement.

D. Practical Applications.

These findings have practical implications for organizations handling large volumes of unstructured data. The joeddav/xlm-roberta-large-xnli model can be used in multilingual document management, while Generative AI can dynamically classify customer queries in service platforms. Integrating these models can improve data processing, decision-making, and productivity.

VI. CONCLUSION

This study highlights the vital role of advanced AI in transforming qualitative data into quantitative insights. The evaluation of joeddav/xlm-roberta-large-xnli, facebook/bart-large-mnli, and Generative AI models demonstrates that Zero-Shot and Generative AI approaches offer reliable, scalable solutions. With high Z-values, joeddav/xlm-roberta-large-xnli excels in multilingual support, while Generative AI shines in adaptability and contextual understanding.

These models enhance data classification accuracy and reduce the need for extensive pre-training, making them invaluable for efficient unstructured data management. Their integration into data systems promises significant improvements in processing accuracy and efficiency. Future research should refine these models and explore new applications, fully harnessing AI's potential in managing unstructured data.

REFERENCES

[1] H.-J. Kong, “Managing unstructured big data in healthcare system,” Healthc. Inform. Res., vol. 25, no. 1, pp. 1–2, 2019.

[2] H. Abburi, M. Suesserman, N. Pudota, B. Veeramani, E. Bowen, and S. Bhattacharya, “Generative AI text classification using ensemble LLM approaches,” in *Proc. Iberian Lang. Eval. Forum*, Jaén, Spain, Sep. 2023. [Online]. Available: https://ceur-ws.org/Vol-3496/autextification-paper14.pdf

[3] C. Han, H. Pei, X. Du, and H. Ji, “Zero-shot classification by logical reasoning on natural language explanations,” in *Proc. 61st Annu. Meeting Assoc. Comput. Linguistics*, Toronto, Canada, Jul. 2023. pp. 8967–8981.

[4] J. Zhang, P. Lertvittayakumjorn, and Y. Guo, “Integrating semantic knowledge to tacklezero-shot text classification,” in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics: Human Lang. Technologies*, Minneapolis, MN, USA, Jun. 2019, pp. 1031–1040.

[5] Forbes Thailand. “Unstructured Data: A treasure trove waiting to be unlocked.” (in Thai), FORBESTHAILAND.com. https://forbesthailand.com/commentaries/Insights/unstructured-data-ขุมทรัพย์ที่รอการปลด (accessed Aug. 1, 2024).

[6] Q. Ye et al., “Zero-shot text classification via reinforced self-training,” in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*, Seattle, WA, USA, Jul. 2020, pp. 3014–3024.

[7] W. Yin, J. Hay, and D. Roth, “Benchmarking zero-shot text classification: Datasets, evaluation and entailment approach,” in *Proc. Conf. Empirical Methods in Natural Lang. Process. and 9th Int. Joint Conf. Natural Lang. Process. (EMNLP-IJCNLP)*, Hong Kong, China, Nov. 2019, pp. 3914–3923.

[8] T. Brown et al., “Language models are few-shot learners,” in *Proc. 34th Conf. Neural Inf. Process. Syst. (NeurIPS)*, Vancouver, Canada, Dec. 2020, pp.1–25.

[9] K. J. Srnka and S. T. Koeszegi, “From words to numbers: How to transform qualitative data into meaningful quantitative results,” *Schmalenbach Bus. Rev.,* vol. 59, pp. 29–57, 2007.

[10] H. Abi Akl, “A ML-LLM pairing for better code comment classification,” in *Proc. 15th Meeting Forum Inf. Retrieval Eval. (FIRE),* Panjim, India, Dec. 2023.

[11] T.-l. Chasupa and W. Paireekreng, “The framework of extracting unstructured usage for big data platform,” in *Proc. 2nd Int. Conf. Big Data Analytics and Pract. (IBDAP)*, Bangkok, Thailand, Aug. 2021, pp. 90–94.

[12] Talance. “Let's take a look! What are the responsibilities of IT positions?.” (in Thai), TALANCE.tech https://www.talance.tech/blog/it-job-responsibility/ (accessed Aug. 1, 2024).

[13] G. Jaimovitch-López, C. Ferri, J. Hernández Orallo, F. Martínez-Plumed, and M. J. Ramírez-Quintana, “Can language models automate data wrangling?,” *Mach. Learn*., vol. 112, pp. 2053–2082, 2023.

[14] H. Ahmad and H. Halim, “Determining sample size for research activities: The case of organizational research," *Selangor Bus. Rev.*, vol. 2, no. 1, pp. 20–34, 2017.

[15] R. Anantha, T. Bethi, D. Vodianik, and S. Chappidi, “Context tuning for retrieval augmented generation,” in *Proc. 1st Workshop on Uncertainty-Aware NLP (UncertaiNLP),* St. Julian’s, Malta, Mar. 2024, pp. 15–22.