Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 257 (2025) 60–67

Procedia Computer Science

www.elsevier.com/locate/procedia

The 16th International Conference on Ambient Systems, Networks and Technologies (ANT)
April 22-24, 2025, Patras, Greece

# LLM-Assisted Qualitative Data Analysis: Security and Privacy Concerns in Gamified Workforce Studies

Aisvarya Adeseye[a,*], Jouni Isoaho[a], Tahir Mohammad[a]

[a] *Department of Computing, University of Turku, Vesilinnantie 5, Turku 20014, Finland*

**Abstract**

Large language models (LLMs) have transformed textual or qualitative data processing and analysis by automating and enhancing interpretive accuracy, particularly in complex areas like cybersecurity, ethics, and compliance. This study examines the effectiveness of local LLMs in analyzing qualitative research using the data gathered from the case study on "perspectives on security and privacy issues associated with the introduction of gamified workforce studies". The research presented in this paper utilized 23 interview transcripts to evaluate three popular LLMs, namely LLaMA, Gemma, and Phi, running on a local infrastructure. We observed that LLaMA focuses on practical data security, Gemma on regulatory compliance, and Phi on ethical transparency and trust-building. By combining these models, researchers can gain a more comprehensive understanding of the complex implications of gamification in workforce studies. Local LLMs provide the added benefit of enhanced data privacy and security by processing sensitive data entirely within a controlled environment. This study explores the system and user prompts that can improve the interpretive accuracy of various qualitative research approaches, such as thematic analysis, frequency analysis, impact level analysis, sensitivity analysis, and disclosure analysis, demonstrating the potential of local LLMs for qualitative analysis for sensitive data. This study recommends the usage of LLMs for the initial stage of the qualitative analysis process to enhance the efficiency and effectiveness of subsequent completely manual or software-assisted manual analysis.



*Keywords:* Large Language Models; Qualitative Data Analysis; Prompt Engineering; Privacy and Security; Gamification; Workforce Studies;

## 1. Introduction

Conventionally, manual coding in qualitative analysis is time-intensive and requires significant effort to categorize and interpret [1]. According to [3], thematic analysis or grounded theory requires careful coding and interpretation to answer research questions. Consequently, analysts often identify recurring patterns, themes and narratives in qualitative data [2]. Alternatively, software tools such as *NVivo*, *Atlas.ti*, and *MAXQDA* are alternatives. However, this

* Aisvarya Adeseye
*E-mail address:* aisvarya.a.adeseye@utu.fi

1877-0509 
10.1016/j.procs.2025.03.011

software requires data to be entered manually because it lacks advanced natural language processing capabilities and cannot automate theme recognition and complex pattern interpretations [4]. However, LLMs can process large textual datasets, identify trends and derive insights through natural language processing (NLP) [5]. They can spot changes in expression and language, which a typical qualitative analysis process would not find. However, despite these strengths, it has not been fully applied to analyze qualitative data. Existing research has mainly focused on the general capacities of LLMs for textual analysis without tailoring them to various qualitative research approaches. Consequently, a structured approach to identify and optimize prompts for diverse qualitative research approaches to improve the interpretive accuracy of LLMs remains notably absent. Online LLMs like ChatGPT and Gemini provide accessible and powerful processing capabilities but raise privacy concerns when dealing with sensitive qualitative data. Also, data processing and storage require external servers raising concerns about data security and privacy compliance [5]. Local LLMs are a viable alternative to mitigate this risk; organizations could use them to localize data processing within their infrastructure. However, the feasibility and performance of local LLMs have not yet been fully tested. This study introduces system and user prompts that could be used in different qualitative research approaches to study the feasibility, performance, and efficiency of using local LLMs for qualitative analysis. This study utilized qualitative data from the *"perspectives on security and privacy issues associated with the introduction of gamified workforce studies"* as a case study. Therefore, this study aims to thematically analyze the qualitative data from the study using a software-assisted manual analysis (baseline analysis with NVivo) and local LLMs analysis (LLaMA, Gemma, Phi). The key research questions (RQ) of the study are:

- **RQ1:** What prompts could be used to optimize the interpretive accuracy of local LLMs for different qualitative research approaches?
- **RQ2:** How feasible, performant, and efficient can local LLMs analyze qualitative data compared to software-assisted manual analysis?

This study explores different aspects of LLM-assisted qualitative analysis, including the technical, ethical, and practical aspects. It offers insights into how researchers and professionals can use this approach to enhance qualitative analysis of sensitive data. The key contribution of this work is the development of optimized prompts that improve precision and relevance when using local LLMs to conduct different qualitative research approaches. Furthermore, this study evaluates the feasibility, efficiency, and performance of local LLMs in qualitative analysis by comparing their effectiveness against the software-assisted manual analysis.

The rest of the paper is organized as follows: Section 2 focuses on the background study of qualitative analysis and LLMs. Section 3 outlines the research methodology used. Section 4 contains the result of the LLM-based qualitative analysis. Section 5 is a comparative overview of the results received from software-assisted manual and LLM-based analysis. Lastly, Section 6 concludes this study.

## 2. Background Study

### *2.1. Qualitative Data Analysis*

Qualitative research analysis systematically analyzes textual data to identify patterns and insights [8]. Literature has introduced various analytical approaches, including thematic, sentiment, disclosure, frequency, and impact analyses. Thematic analysis identifies recurrent patterns and themes in textual data, enabling a structured comprehensive participants' perspectives [9]. Sentiment Analysis assesses the emotional tone, capturing participants' attitudes, opinions, or sentiments [9]. Disclosure Analysis examines the patterns of information shared and withheld, offering insights into the communication practices of the participants [9]. Frequency Analysis quantifies key terms to highlight dominant ideas in the discussion [10], while impact Analysis evaluates how specific themes influence the discussions, providing practical implications for research findings [10]. Qualitative analysis is conducted through two main methods: manual coding and software-assisted analysis. Although manual coding offers deep interpretive insights, it is time-consuming and prone to bias [11]. Software-assisted tools such as *NVivo*, *MAXQDA*, and *Atlas.Ti* enhances efficiency and structure but relies heavily on significant manual effort for coding and theme recognition [12] [13]. This limitation highlights the need for advanced text analysis technologies to conduct complex thematic analysis.

### 2.2. Large Language Models (LLMs)

LLMs leverage deep learning techniques to understand context, recognize patterns [14], and generate meaningful outputs from complex and voluminous textual datasets. Their sophisticated text generation and pattern recognition skills enhance the depth and objectivity of text analysis, minimizing the need for extensive researcher involvement [15].

LLMs can be deployed either online or locally. Online LLMs such as ChatGPT, Copilot, Gemini, DeepSeek AI are hosted on external servers and accessed via the Internet. In contrast, local LLMs like LLaMA, Gemma, and Phi can be deployed in-house and accessed via a private network. Local LLMs are preferable for sensitive analysis due to improved security and privacy. Unlike online LLMs, they keep their data within a controlled environment, minimizing data leakage risk while ensuring better compliance with data protection requirements. However, local LLMs require extensive customization and in-house expertise for maintenance and are limited by hardware capabilities. Although online models operate on a pay-as-you-go basis, local LLMs require higher initial setup costs but eliminate long-term reliance on external providers.

LLMs process textual data via system and user prompts. System prompts are predefined instructions given to the model to make it apply instructions uniformly and consistently [17]. On the contrary, user prompts are dynamic instructions given to the model to extract specific insight and facilitate flexible data interpretation. Furthermore, prompts are vulnerable to extraction attacks using techniques like output2prompt, raising privacy and security concerns, especially with online LLMs [17]. Local LLMs mitigate this risk by keeping data in a controlled environment, e.g., an office network. For qualitative analysis to be effective, it is important to balance the system and user prompts because a well-crafted system prompt enhances consistency. In contrast, user prompt enhances adaptability, which is necessary to ensure accurate LLM-assisted qualitative analysis.

## 3. Methodology

**Case Study:** Traditionally, workforce studies rely on surveys and interviews, which often suffer from low engagement and survey fatigue, leading to poor response rates [6]. To address this, an extensive feasibility study of using gamification, i.e. the introduction of game-like elements [7] to enhance participation in workforce studies, was conducted. The study gathers *“diverse perspectives on potential security and privacy concerns associated with the introduction of gamified workforce studies”*. To investigate these concerns, semi-structured interviews were conducted by the authors with 23 experts from diverse organizations such as NGOs, companies, and universities. The participants answered the same open-ended questions. Each interview lasted for between 45 to 60 minutes. The gathered insights ensured a comprehensive understanding of the potential security and privacy concerns of gamified workforce studies.

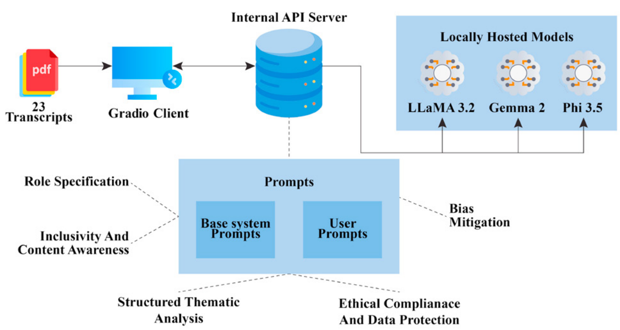


Fig. 1. Experiment Setting of LLMs for Qualitative Analysis.

**Experiment Setting:**A total of 23 transcripts were used for the analysis. Each transcript contained between 8,000 -13,000 words. The study utilizes LLaMA (Meta) v3.2 (1B parameters), Gemma (Google) v2 (2B parameters), and Phi (Microsoft) v3.5 (3.8B parameters), deployed on a system configured with an 8-core CPU, 32GB RAM, and a 500GB SSD for model evaluation. These LLMs were selected due to diverse architectural optimizations, and broad adoption in NLP tasks, enabling a balanced assessment of performance, efficiency, and feasibility for qualitative data analysis. Fig 1 shows that the Gradio client functions as a user interface for users to input the transcripts, and the internal API server functions as a central hub for routing between the client and models and also functions as a preprocessing and post-processing of data.

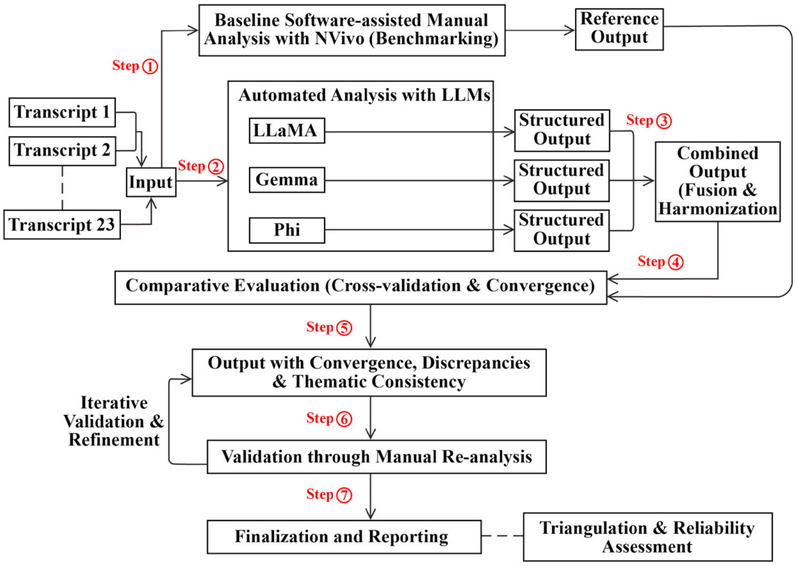


Fig. 2. Steps involved in studying the performance evaluation of the LLMs for qualitative analysis.

**Experiment Stages:** The detailed step to perform this study's benchmarking and LLMs performance evaluation can be seen in fig 2.

**Prompt Design:** The strategy to determine the prompt was to combine symbolic reasoning [16] for identifying keywords, conjunctions, and keyword arrangement, along with iterative prompting [16] to refine the prompt based on output quality.

Additionally, the base system prompt was designed with key factors in mind: *Role Specification*, ensuring the LLM functions as a research analyst; *Structured Thematic Analysis*, guiding it to extract patterns and key themes; *Inclusivity and Context Awareness*, directing it to consider cultural, linguistic, and contextual variations; *Bias Mitigation*, ensuring neutrality in interpretation; and *Ethical Compliance and Data Protection*, safeguarding confidentiality by preventing the inclusion of personally identifiable information. This structured approach enhances prompt effectiveness, interpretive accuracy, and ethical reliability in LLM-assisted qualitative analysis. Neglecting these factors reduces the reliability and ethical integrity of the qualitative analysis, making the findings less credible and potentially problematic in sensitive research contexts. The finalized prompts that were used across the models are provided below.

*"You are a research analyst* ***(Role Specification)*** *tasked with analyzing transcripts from interviews with participants from diverse cultural, demographic, and industrial backgrounds. Your job is to identify insights, patterns, and key themes* ***(Structured Thematic Analysis)*** *in the transcripts that I will provide. Also, ensure that cultural, linguistic, and contextual distinctions are considered* ***(Inclusivity and Context Awareness)*** *to reflect everyone's perspective accurately. Maintain neutrality and avoid favoring any particular viewpoint* ***(Bias Mitigation)****. Lastly, ensure that you adhere to ethical guidelines and data protection standards* ***(Ethical Compliance and Data Protection)****, such as avoiding personally identifiable information that could jeopardize confidentiality and data integrity."*

The table 1 outlines the user prompts used to interact with the LLMs to perform various qualitative analysis tasks. The user prompts were designed to include only the relevant factors, ensuring a structured and targeted prompt.

## 4. Results Analysis

The following sections analyze the results from the five user prompts from the three local LLMs: LLaMA, Gemma, and Phi.

### *4.1. Contextual Overview*

In qualitative research, defining the interview context is essential for the researcher as it shapes the interpretation of the participant's responses. For LLMs, context prompts provide an understanding of the interpretative lens through which the analysis is conducted, which enables researchers to align or adjust the model's perspective as required. A partial view of the results received from the models and the interpretation focus can be seen in Table 2. The result shows that each local LLM interpreted the context with a particular focus. For example, LLaMA's conversation prioritized security-focused analysis, Gemma aligned with regulatory compliance, and Phi offered insight into ethical

Table 1. Finalized user Prompts utlized to conduct diverse qualitative analysis task

| No. | Tasks | User Prompts |
|---|---|---|
| 1 | Context Overview | *"What is the context? Include the main topics discussed by all backgrounds of participants **(Inclusivity and Context Awareness)**."* |
| 2 | Key Areas and respective Thematic, Importance, and Frequency Analysis | *"Identify the key areas of this transcript and assess their relative importance based on participant emphasis and thematic relevance **(Structured Thematic Analysis)**. Generate a table listing the identified key areas, corresponding themes, thematic coding, the frequency of each code, and the overall occurrence percentage for each key area as importance **(Symbolic Reasoning for Data Structuring)**."* |
| 3 | Summary Findings | *"List the key findings for the areas you previously identified. Structure the summary in clear sentences, categorizing findings by theme and importance **(Role Specification for Analytical Summarization)**."* |
| 4 | Sentiment Analysis | *"Analyse participants' feelings about the key areas identified in the transcript. Provide a sentiment classification (e.g., positive, negative, neutral) **(Bias Mitigation and Context Sensitivity)**."* |
| 5 | Disclosure Analysis | *"Perform a disclosure analysis on key areas. Specify the nature of disclosures (explicit vs. implicit), and assess any patterns in disclosure behavior **(Ethical Compliance and Data Protection)**."* |

Table 2. Partial context output of the LLM Models' and Interpretation Focus

| LLM Models | Extracted Context Output | Interpretation Focus |
|---|---|---|
| LLaMA | *"...there is a need to balance benefits with secure data handling. Raises concerns about stakeholder-wide data access and unauthorized access risks ..."* | Security Risks |
| Gemma | *"...the importance of strict access controls and ensuring GDPR compliance in gamified platforms. Highlights the role of compliance teams in maintaining legal adherence ..."* | Regulatory Compliance |
| Phi | *"...transparent communication about data collection, the need to engender employee trust, and strategies to increase comfort with workplace gamification ..."* | Ethical Transparency |

transparency. This difference in the output underscores the importance of selecting an LLM-based research context to ensure alignment with the qualitative analysis.

### *4.2. Key Areas, Importance Analysis, Thematic Coding Analysis and Frequency Analysis*

This is the key aspect of qualitative analysis, where the LLMs identify the key areas of the data and their importance level. Then, it identifies codes to capture all the sub-areas and determine the frequency of use in the transcripts. Table 3 shows that LLaMA uses practical or actionable approaches by identifying specific data security and privacy risks while emphasizing gamification's benefits and risks. This indicates a balanced focus on concrete strategies to balance the advantages of gamification with secure data handling. For instance, LLaMA focuses on security and engagement benefits, frequently coding for *"Access Restrictions"* and *"Employee Motivation"*. Gemma strongly emphasizes regulatory compliance, reflecting its technical and compliance-driven approach to cybersecurity. For example, *"GDPR Adherence"* and *"Data Subject Rights"*. Phi focuses on three key points: the ethical implications, building employee trust and transparency, and this can be seen in themes like *"Data Usage Disclosure"* and *"Equitable Data Treatment"*.

### *4.3. Key Findings Summary*

This provides a solid foundation for deeper analysis and accurate reporting by precisely capturing key insights. LLaMA's priority is practical or actionable security aspects. *"Specific access restrictions to protect sensitive employee data"* and *"Firewall protections can prevent unauthorized breaches in gamified systems"* were some of the observed findings. *"Employee motivation can be enhanced through structured learning incentives"* was another finding on gamification benefits. LLaMA focused more on practicality, which results in less emphasis on more profound ethical implications. Gemma's regulatory perspective is detailed. For example, one of its findings was about how *"ensuring GDPR adherence is fundamental for handling employee data legally"*. It also stresses the importance of *"setting clear data retention policies"* to comply with legal standards. Gemma also suggests that *"authorization protocols are necessary to manage data access within a gamified system"*. These findings are structured as key takeaways covering

Table 3. Analysis of Key Areas and their Impact on LLMs

| Key Area | Importance | Themes | Coding for Themes | Frequency (out of 23) |
|---|---|---|---|---|
| Data Security and Privacy Risks | LLaMA 40%, Gemma 20%, Phi 15% | Unauthorized Access | ”Access Restrictions” | LLaMA: 10, Gemma: 6 |
| | | | ”Firewall Protections” | LLaMA: 5, Gemma: 4 |
| | | Data Encryption | ”End-to-End Encryption” | LLaMA: 7, Gemma: 5 |
| | | | ”Data Masking” | LLaMA: 3, Gemma: 2 |
| | | | ”Data Encryption Standards” | Gemma: 3 |
| | | Anonymization of Employee Data | ”Pseudonymization” | LLaMA: 5, Phi: 3 |
| | | | ”Data Aggregation” | LLaMA: 2, Phi: 2 |
| | | | ”Anonymized Reporting” | Phi: 3 |
| | | Risk Mitigation Techniques | ”Regular Security Audits” | LLaMA: 6 |
| | | | ”Data Minimization” | LLaMA: 4 |
| | | | ”Employee Access Control” | LLaMA: 2 |
| Regulatory Compliance | LLaMA 5%, Gemma 45% | GDPR and Legal Standards | ”GDPR Adherence” | Gemma: 9 |
| | | | ”Data Subject Rights” | Gemma: 6 |
| | | | ”Data Retention Policies” | Gemma: 3 |
| | | Role-based Access Control | ”Access Levels” | Gemma: 6 |
| | | | ”Authorization Protocols” | Gemma: 4 |
| | | | ”User Privilege Management” | Gemma: 2 |
| | | Data Auditing and Monitoring | ”Regular Data Audits” | LLaMA: 4, Gemma: 4 |
| | | | ”Anomaly Detection” | LLaMA: 2, Gemma: 2 |
| | | | ”Data Integrity Checks” | LLaMA: 2, Gemma: 2 |
| Employee Trust and Transparency | LLaMA 10%, PHI 30% | Transparency in Data Collection | ”Clear Communication Policies” | LLaMA: 5, Phi: 6 |
| | | | ”Data Usage Disclosure” | Phi: 8 |
| | | | ”Purpose of Data Collection” | Phi: 5 |
| | | Communication of Data Policies | ”Data Policy Updates” | Phi: 5 |
| | | | ”Employee Consent” | LLaMA: 3, Phi: 3 |
| | | | ”Right to Withdraw Participation” | Phi: 2 |
| | | Ensuring Voluntary Participation | ”Opt-In/Opt-Out Options” | Phi: 3 |
| | | | ”Consent Management” | Phi: 2 |
| | | | ”Voluntary Participation Statements” | LLaMA: 1, Phi: 1 |
| Ethical Implications | LLaMA 15%, Gemma 10%, Phi 35% | Fairness in Data Use | ”Bias Mitigation” | LLaMA: 4, Gemma: 3, Phi: 6 |
| | | | ”Equitable Data Treatment” | Phi: 5 |
| | | | ”Non-Discriminatory Practices” | Phi: 3 |
| | | Consent and Autonomy | ”Informed Consent” | LLaMA: 3, Phi: 5 |
| | | | ”Data Autonomy” | LLaMA: 4, Gemma: 4, Phi: 4 |
| | | | ”Opt-Out Freedom” | Phi: 3 |
| | | Employee Well-being and Respect | ”Non-Intrusive Monitoring” | Phi: 4 |
| | | | ”Respect for Privacy” | LLaMA: 4, Gemma: 4, Phi: 3 |
| | | | ”Employee Dignity” | Phi: 1 |

the necessary legal compliance. phasizes employee trust and ethical implications. For example, it emphasizes that *“transparent data usage disclosure builds trust when employees are informed how their data is handled”*. It stresses *“equitable treatment in data practices to avoid bias”*. Phi also suggests that *“employees should have the option to opt-out to respect personal data autonomy”*. The findings were extensive and detailed, presented as specific ethical points that deepen the understanding of relational and trust-building practices.

### *4.4. Sentiment Analysis*

Sentiment analysis systematically assesses the emotional tone of the key data areas and groups complex qualitative expressions into obvious sentiment indicators. This approach provides researchers with complete background knowledge of attitudes and emotions. LLaMA analyzes the participant’s emotional responses in detail, covering their concerns about gamification. For example, LLaMA observed that participants *“feel reassured by specific access restrictions”*. This observation implies that safety awareness is enhanced when security measures are implemented. Despite this, feelings of apprehension were noted, and participants confirmed that *“firewall protections alone may not address all vulnerabilities”*. This implies a residual concern for security inadequacy. Gemma is moderate in its view because it focuses more on the participants' feelings based on their trust and compliance with regulatory measures. It shows that participants *"feel confident when data handling aligns with GDPR"*, indicating reassurance that comes with legal adherence. Gemma also factors in the participants' level of concern. Hence, it reported that participants feel *"uncomfortable if data retention policies are unclear"*. This is an indication of their anxiety because of misuse. Phi presents a concise analysis focusing on trust, ethical concerns, and participants' relational feelings toward gamification. Participants feel *"valued when data usage is transparent"*, indicating the emotional significance of open data practices. Phi reveals some concerns; participants *"feel uneasy if they cannot opt-out"*, suggesting that autonomy loss triggers discomfort.

### *4.5. Disclosure Analysis*

Analysis of disclosure systematically allows researchers to examine information dissemination level and nature in critical data areas. Such a structure is crucial because it compartmentalizes complex data into specific disclosure trends. It also lets the researchers have a fair understanding of shared information's openness, sensitivity, and relevance. LLaMA provides a detailed analysis of the level of disclosure across key areas, specifying what information is shared with employees and how. For example, LLaMA notes, *"Access restrictions are disclosed fully, allowing employees to understand how their data is secured"*. It also highlights partial disclosures, where participants report that *"firewall protections are mentioned but not in detail"*, suggesting that security practices are disclosed partially and fully. Similarly, Gemma's disclosure levels focus on regulatory specifications, noting disclosures on compliance. For example, *"GDPR and data retention policies are disclosed to meet legal standards"* This shows mandatory adherence to compliance information. Gemma's assessment cuts across complete and partial disclosure, adding that *"authorization protocols are not consistently disclosed"*, which shows an existing gap in access control disclosure. Phi's focus is mainly on addressing employee concerns regarding data security. Phi observes that *"Employees are concerned about data security"* and highlights a unique issue: a perceived lack of transparency in certain areas. Specifically, Phi identified that *"employees feel uneasy when there is no explanation of how their data is used"*. This suggests that Phi not only recognizes the technical aspects of data security but also emphasizes the emotional and trust-related implications of insufficient disclosure.

## 5. Evaluation and Discussion

The findings reveal that neither software-assisted manual analysis with NVivo nor LLM-based analysis could identify all key areas and coding in the qualitative data. Compared to the finalized validation results in table 4, software-assisted manual analysis with NVivo and LLM-based analysis did not always achieve the perfect score for the identified coding. Both analyses performed worst for employee trust and transparency due to difficulty identifying implicit coding and inconsistent keyword matching. In Table 4, each method has its strengths and limitations, enforcing the need for a combined, iterative approach in qualitative analysis. The significant advantage of LLMs was the speed of analysis as they completed the tasks in under 8 minutes, while software-assisted manual analysis using NVivo took approximately a day (24 hours), making them effective for large datasets. Also, the model output varies, requiring a harmonization process to ensure consistency. It is observed that the quality of the output from LLMs relies on prompting techniques, training data and model-specific biases. Therefore, while LLMs serve as valuable tools for accelerating qualitative analysis, their output should not be solely relied upon but used as a starting point for further analysis. Also, running the models on a Graphics Processing Unit (GPU) will reduce the LLM's computation time from minutes to seconds. The findings suggest that LLMs should be considered complementary tools rather than replacements for traditional qualitative analysis methods.

## 6. Conclusion

The evaluation demonstrates the unique strengths of LLaMA.2, Gemma2 and Phi3.5 in performing qualitative analysis within the context of privacy and security. LLaMA excels in detailed security insights, Gemma focuses on compliance, and Phi on ethical transparency, creating a robust analytical framework when used together. Integrating LLMs with GPU acceleration enables timely and detailed qualitative analysis. This research utilized two types of

Table 4. Comparative observation of coding extracted for key areas across software-assisted manual analysis with NVivo, LLMs and validated results.

| Key Areas | No of Coding Identified | | | | |
|---|---|---|---|---|---|
| | Software-assisted Manual Analysis (Benchmarking) | LLMs | | | Finalized Validation |
| | | LLaMA | Gemma | Phi | |
| Data Security and Privacy Risks | 10 | 9 | 5 | 3 | 11 |
| Regulatory Compliance | 9 | 3 | 9 | 0 | 9 |
| Employee Trust and Transparency | 7 | 3 | 0 | 9 | 12 |
| Ethical Implications | 8 | 4 | 3 | 9 | 9 |

prompts: base system and user prompts. The output results of the user prompt depend on the base system prompt, making it necessary to have a tailored system prompt for the case study. Notably, the accuracy of the user prompts is influenced by keyword selection, keyword sequence, and model size. Therefore, it is essential to prioritize the arrangement of keywords based on their relevance and importance level. LLMs outperformed software-assisted manual analysis in processing time; however, their results are based mainly on the context set by the researcher. While LLMs identified some key areas overlooked by manual analysis, they also failed to identify areas manual analysis could find. Therefore, this study suggests that LLMs should be used as a preliminary step to enhance the efficiency of traditional qualitative analysis methods. Future applications of LLMs could extend to the implementation of gamified elements and the automation of workforce study reporting.

## References


[1] Parameswaran, U. D., Ozawa-Kirk, J. L., and Latendresse, G. (2020) "To live (code) or to not: A new method for coding in qualitative research," *Qualitative Social Work*, vol. 19, no. 4, pp. 630–644.

[2] Hadi, M. U., Qureshi, R., Shah, A., Irfan, M., Zafar, A., Shaikh, M. B., Akhtar, N., Wu, J., and Mirjalili, S. (2023) "A survey on large language models: Applications, challenges, limitations, and practical usage," *Authorea Preprints*.

[3] Finlay, L. (2021) "Thematic analysis:The 'good',the 'bad' and the 'ugly'," *European Journal for Qualitative Research in Psychotherapy*, vol. 11, pp. 103–116.

[4] Wright, S. (2021) *Automating Insights. Analysing Student Feedback in Higher Education: Using Text-Mining to Interpret the Student Voice*.

[5] Khan, W., Daud, A., Khan, K., Muhammad, S., and Haq, R. (2023) "Exploring the frontiers of deep learning and natural language processing: A comprehensive overview of key challenges and emerging trends," *Natural Language Processing Journal*, vol. 100026.

[6] Salleh, N., Hoda, R., Su, M. T., Kanij, T., and Grundy, J. (2018) "Recruitment, engagement and feedback in empirical software engineering studies in industrial contexts," *Information and Software Technology*, vol. 98, pp. 161–172.

[7] Schöbel, S.M., Janson, A., and Söllner, M. (2020) "Capturing the complexity of gamification elements: A holistic approach for analysing existing and deriving novel gamification designs," *European Journal of Information Systems*, vol. 29, no. 6, pp. 641–668.

[8] Kuckartz, U. (2014) *Qualitative text analysis: A guide to methods, practice and using software*, Sage.

[9] Jaspal, R. (2020) "Content analysis, thematic analysis and discourse analysis," *Research Methods in Psychology*, vol. 1, pp. 285–312.

[10] Li, Y., and Zhang, S. (2022) "Qualitative data analysis," in *Applied Research Methods in Urban and Regional Planning*, Cham: Springer International Publishing, pp. 149–165.

[11] Belotto, M. J. (2018) "Data analysis methods for qualitative research: Managing the challenges of coding, interrater reliability, and thematic analysis," *The Qualitative Report*, vol. 23, no. 11, pp. 2622–2633.

[12] Oliveira, M., Bitencourt, C., Teixeira, E., and Santos, A. C. (2013) "Thematic content analysis: Is there a difference between the support provided by the MAXQDA® and NVivo® software packages," in Joanne Evans and Lester Hunt (eds) *Proc. 12th European Conf. Research Methods for Business and Management Studies*, pp. 304–314.

[13] Yu, W., Zhu, C., Li, Z., Hu, Z., Wang, Q., Ji, H., and Jiang, M. (2022) "A survey of knowledge-enhanced text generation," *ACM Computing Surveys*, vol. 54, no. 11s, pp. 1–38.

[14] Le, D., Indurkhya, B., and Nakagawa, M. (2019) "Pattern generation strategies for improving recognition of handwritten mathematical expressions," *Pattern Recognition Letters*, vol. 128, pp. 255–262.

[15] Dashtipour, K., Poria, S., Hussain, A., Cambria, E., Hawalah, A. Y., Gelbukh, A., and Zhou, Q. (2016) "Multilingual sentiment analysis: State of the art and independent comparison of techniques," *Cognitive Computation*, vol. 8, pp. 757–771.

[16] Shah, C. (2024) "From Prompt Engineering to Prompt Science With Human in the Loop," *arXiv preprint arXiv:2401.04122*.

[17] Zhang, C., Morris, J. X., and Shmatikov, V. (2024) "Extracting Prompts by Inverting LLM Outputs," *arXiv preprint arXiv:2405.15012v2*.