

Analyzing Patient Complaints in Online Reviews of Private Hospitals in Selangor, Malaysia: A Study Using LLM-Assisted Content Analysis (LACA)

Muhammad Hafiz Sulaiman, Nora Muda, Fatimah Abdul Razak

Submitted to: JMIR Formative Research
on: November 21, 2024

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript.....	5
Supplementary Files.....	50
Figures	51
Figure 1.....	52
Figure 2.....	53
Figure 3.....	54
Figure 4.....	55
Multimedia Appendixes	56
Multimedia Appendix 1.....	57
Multimedia Appendix 2.....	57
Multimedia Appendix 3.....	57
Multimedia Appendix 4.....	57
Multimedia Appendix 5.....	57
Multimedia Appendix 6.....	57
Multimedia Appendix 7.....	57

Analyzing Patient Complaints in Online Reviews of Private Hospitals in Selangor, Malaysia: A Study Using LLM-Assisted Content Analysis (LACA)

Muhammad Hafiz Sulaiman^{1, 2*} MD, MS; Nora Muda^{1*} PhD; Fatimah Abdul Razak^{1*} PhD

¹Department of Mathematical Sciences Faculty of Science and Technology National University of Malaysia Bangi MY

²Quality Unit Clinical Management Section Hospital Sultan Idris Shah Serdang MY

*these authors contributed equally

Corresponding Author:

Muhammad Hafiz Sulaiman MD, MS
Department of Mathematical Sciences
Faculty of Science and Technology
National University of Malaysia
Lingkungan Ilmu
Bangi
MY

Abstract

Background: LLM-Assisted Content Analysis (LACA) is a modification of traditional content analysis, leveraging the Large Language Model (LLM) to co-develop codebooks and automatically assign thematic codes to online reviews dataset.

Objective: This study aims to develop and validate the use of LACA for analyzing hospital online reviews, and to identify themes of issues from online reviews using this method.

Methods: Online reviews for 53 private hospitals in Selangor Malaysia were acquired. Fake reviews were filtered out using natural language processing and machine learning algorithms trained on yelp.com validated datasets. gpt-4o-mini model API was then applied to filter out reviews without any quality issue. 200 of the remaining reviews were randomly extracted and fed into gpt-4o-mini model API to produce a codebook validated through parallel human-LLM coding to establish inter-rater reliability. The codebook was then used to code (label) all reviews in the dataset. The thematic codes were then summarized into themes using factor analysis to increase interpretability.

Results: 14,938 online reviews were acquired in which 1,121 were fake, 1,279 contain negative sentiments (12%) and 9,635 do not contain any negative sentiment (88%). gpt-4o-mini model subsequently inducted 41 thematic codes together with their definitions. Average Human-GPT inter-rater reliability is perfect ($\kappa = 0.81$). Factor analysis identified six interpretable latent factors: 'Service & Communication Effectiveness', 'Clinical Care & Patient Experience', 'Facilities & Amenities Quality', 'Appointment & Patient Flow', 'Financial & Insurance Management' and 'Patient Rights & Accessibility'. The cumulative explained variance for the six factors is 0.74 and Cronbach's alpha is between 0.88 - 0.97 (good - excellent) for all factors except factor 6 (0.61 - questionable). The factors identified follow a global pattern of issues identified from literature.

Conclusions: A data collection and processing pipeline consisting of Python Selenium, the gpt-4o-mini model API, and a factor analysis module can support valid and reliable thematic analysis. Despite the potential for collection and information bias in online reviews, LACA of online reviews is cost-effective, time-efficient, and can be performed in real time, helping hospital managers develop hypotheses for further investigations promptly.

(JMIR Preprints 21/11/2024:69075)

DOI: <https://doi.org/10.2196/preprints.69075>

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <http://www.jmir.org/>

No. Please do not make my accepted manuscript PDF available to anyone.



Original Manuscript

Original Paper

Dr. Muhammad Hafiz Sulaiman, MD, MS^{1,2}

Associate Prof. Dr. Nora Muda, PhD¹

Associate Prof. Dr. Fatimah Abdul Razak, PhD¹

1. Department of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, Bangi, Malaysia
2. Hospital Sultan Idris Shah, Serdang, Selangor, Malaysia

Analyzing Patient Complaints in Online Reviews of Private Hospitals in Selangor, Malaysia: A Study Using LLM-Assisted Content Analysis (LACA)

Abstract

Background: LLM-Assisted Content Analysis (LACA) is a modification of traditional content analysis, leveraging the Large Language Model (LLM) to co-develop codebooks and automatically assign thematic codes to online reviews dataset. **Objectives:** This study aims to develop and validate the use of LACA for analyzing hospital online reviews, and to identify themes of issues from online reviews using this method. **Method:** Online reviews for 53 private hospitals in Selangor Malaysia were acquired. Fake reviews were filtered out using natural language processing and machine learning algorithms trained on yelp.com validated datasets. gpt-4o-mini model API was then applied to filter out reviews without any quality issue. 200 of the remaining reviews were randomly extracted and fed into gpt-4o-mini model API to produce a codebook validated through parallel human-LLM coding to establish inter-rater reliability. The codebook was then used to code (label) all reviews in the dataset. The thematic codes were then summarized into themes using factor analysis to increase interpretability. **Results:** 14,938 online reviews were acquired in which 1,121 were fake, 1,279 contain negative sentiments (12%) and 9,635 do not contain any negative sentiment (88%). gpt-4o-mini model subsequently inducted 41 thematic codes together with their definitions. Average Human-GPT inter-rater reliability is perfect (kappa = 0.81). Factor analysis identified six interpretable latent factors: 'Service & Communication Effectiveness', 'Clinical Care & Patient Experience', 'Facilities & Amenities Quality', 'Appointment & Patient Flow', 'Financial & Insurance Management' and 'Patient Rights & Accessibility'. The cumulative explained variance for the six factors is 0.74 and Cronbach's alpha is between 0.88 - 0.97 (good - excellent) for all factors except factor 6 (0.61 - questionable). The factors identified follow a global pattern of issues identified from literature. **Conclusion:** A data collection and processing pipeline consisting of Python Selenium, the gpt-4o-mini model API, and a factor analysis module can support valid and reliable thematic analysis. Despite the potential for collection and information bias in online reviews, LACA of online reviews is cost-effective, time-efficient, and can be performed in real time, helping hospital managers develop hypotheses for further investigations promptly.

Keywords: Large Language Model; Hospital Quality; Patient Satisfaction; Big Data; Online Review.

Introduction

Quality Improvement Activities in Hospitals

Getting feedback from patients and families is important to continuously improve patient care and ensuring patients and families satisfaction within healthcare settings [^{1,2}]. Patients' and families' satisfaction is crucial for repeat visits and economic sustainability of the healthcare provider therefore, management teams need to empathize with their patients and families' understandings, feelings, and behaviors in order to thrive in a competitive healthcare market.

Online Reviews

The use of traditional survey like SERVQUAL to measure quality in Malaysian Healthcare is documented by [^{3,4,5}]. Observations, formal interviews, and surveys such as SERVQUAL are standard methods for collecting feedback from patients and families, but all require a significant amount of time for data collection [^{6,7}]. Since these methods are resource-intensive, obtaining unbiased results will require researchers to invest more time and money, otherwise the number of respondents is limited which could lead to low study power and false negative results. Additionally, ethnographic studies often fall short in their ability to observe patients before, during and after hospital stay due to privacy concerns both from clinicians' side and patients' side. [⁸]

Use of online reviews to gather feedback on patients' experiences and opinions can help hospital managers address the limitations mentioned above. Unlike ethnographic studies and interviews, online reviews by patients and families have no spatial limitations, meaning that patients and families can share their feelings, experiences, and opinions throughout their entire journey—before, during, and after their hospital stay or visit. Online reviews are also readily available on the internet, and with the aid of LLM-assisted content analysis, we can include a larger number of respondents, thereby reducing the risk of Type II error (false negatives) caused by an insufficient sample size.

[⁹] mentioned the advantages of using online reviews for hospital quality improvements including the diversity of domains reported in online reviews. Traditional surveys such as HCAHPS have fixed domains in which the questions are based on and have origins derive from 1995. Since patients' indications and experience for hospitalization has changed greatly, using a fixed set of surveys would set a barrier to fully understanding current patients' needs. His study found 12 online review domains not otherwise reflected in HCAHPS.

[¹⁰] in his article however raised the need of healthcare organizations to change in accordance with industrial revolution 4.0 by using online reviews to understand patients' and families' interest, desires and values. He too mentioned that the use of traditional surveys like HCAHPS and SERVQUAL are restrictive in the ways that these surveys are fixed, time intensive, lengthy, fail to identify the causes of concern, and subjected to response and selection bias. The author suggested utilizing online reviews such as Facebook online reviews on hospitals' pages as new sources for quality monitoring in hospitals and using supervised machine learning to train machine learning models to classify these reviews into SERVQUAL domains.

Publicly Available Online Data

The exponential growth of digital communication channels has transformed healthcare feedback mechanisms in Malaysia and globally. Malaysia's internet adoption rate reached 96.8% in 2023 [¹¹]

with social media users aged 18 and above exceeding 24.8 million, representing 99.8% of the population [12]. This digital transformation has generated substantial healthcare-related user-generated content.

While global data usage was projected to reach 35.2ZB by 2020 (Cai et al., 2018), Malaysia's healthcare sector has seen its own surge in digital footprint, with an estimated 54.7% of all patients used the internet to search for health information [13] and 91.2% healthcare workers have good eHealth literacy [14]. Online data reflecting patients' experiences in Malaysian healthcare settings are increasingly available through various digital channels, including social media platforms, hospital review websites, healthcare forums, and patient blogs. The Health White Paper for Malaysia further emphasizes the importance of leveraging these digital data sources for healthcare quality improvement [15].

Study Designs

Our study is exploratory in nature and our research questions are: **(1)** How to apply LACA satisfactorily on hospital online reviews? **(2)** What are the themes of issues identified by LACA on hospital online reviews? The purposes of this study are **(1)** to develop and recommend a method for analyzing hospital online reviews to serve as an alternative to direct observations or interviews on patients and families in hospital settings. **(2)** To then identify themes of current issues in private hospitals in Selangor, Malaysia. For these purposes, we developed three hypotheses: **(1)** LACA developed based on our methods produces satisfactory coding works equivalent to a human coder with Cohen's Kappa > 0.80, **(2)** The themes identified from factor analysis produces Cronbach's alpha of > 0.70 on all factors with interpretable items.

The population of study is patients or families who have posted their reviews online on private hospitals in the state of Selangor, Malaysia from 1st January 2023 to 31st December 2023. All 53 private hospitals in the state were included in this study. This population was chosen because researchers and experts involved in this study have good knowledge of local private hospitals as compared to hospitals located somewhere else, therefore, enabling them to contribute to the qualitative inputs needed in this research. This study includes all reviews posted online for all private hospitals inside Selangor using universal sampling so that we could get as much diversity of reviews as possible, a point of advantage over traditional observations or interviews.

The online reviews are then filtered to exclude reviews that are not accompanied by any comments. Since online reviews are subjected to manipulation by hospitals, we also exclude fake reviews using natural language processing and machine learning algorithms to make sure that the results represent the real patients' suggestions, opinions and experiences. Detailed explanations on this method are discussed below.

Methodology

Thematic Analysis

Thematic analysis is a qualitative research method used to identify, analyze, and report patterns or themes within data. It begins with researchers immersing themselves in the data to gain a deep understanding, which involves repeatedly reading and reviewing the material. Following this, they generate initial codes by systematically tagging relevant sections of the data with short labels that

capture key aspects. These codes are then organized into potential themes—broader patterns that reflect significant features related to the research question.

Researchers review and refine the themes created to ensure they accurately represent the data and fit together cohesively. Each theme is clearly defined and named, and the final step involves writing up the findings to provide a comprehensive interpretation of the data, weaving together the themes to offer insightful conclusions. Thematic analysis is valued for its flexibility and ability to uncover patterns within complex qualitative data. Most of the thematic analysis framework used in this study is based [16].

Inductive Coding in Thematic Analysis

Inductive coding in thematic analysis is a process where researchers develop codes directly from the data, rather than applying predetermined categories or theoretical frameworks (which applies in deductive coding). It begins with a thorough examination of the data to gain an in-depth understanding. Researchers then identify significant segments of text and create codes based on the content and meaning of these segments. These initial codes are descriptive and reflect the language and concepts used by the participants.

As coding progresses, similar codes are grouped together to form broader themes, which emerge naturally from the data itself. This approach allows for a more grounded analysis, as themes are developed from the participants' perspectives rather than imposed by external theories. The themes are then reviewed and refined to ensure they accurately represent the data and provide a coherent interpretation. Inductive coding is particularly useful for exploring new research areas and gaining insights that are deeply rooted in the data [17].

Large Language Model (LLM)

Traditional methods of analyzing text data from surveys and reviews often pose significant challenges, including time-consuming manual processes and resource-intensive endeavors [18]. Moreover, the sheer volume and unstructured nature of textual data available on the internet further exacerbate the complexity of analyzing and extracting actionable insights using conventional methodologies. Given the overwhelming amount of data deriving from online platforms, attention is increasingly turning towards automated content analysis instead of pure qualitative content analysis [19].

With the emergence of Large Language Models (LLMs), such as gpt-4o-mini model, healthcare institutions now possess a powerful tool to navigate and extract valuable insights from the vast expanse of unstructured text data available online. [20] confirms the fact that text mining in big data analytics is emerging as a powerful tool for harnessing the power of unstructured textual data by analyzing it to extract new knowledge and to identify significant patterns and correlations hidden in the data.

LLMs are sophisticated artificial intelligence models trained on large corpora of text data [21], enabling them to understand and generate human-like language with remarkable accuracy and fluency. GPT architecture incorporates attention mechanisms and feed-forward neural networks to predict the next word in a sequence, which improved LLM functionality significantly. [22] Leveraging advanced natural language processing (NLP) techniques, LLMs excel in tasks such as text summarization, thematic analysis [23] and sentiment analysis [24], making them well-suited for analyzing qualitative data in healthcare contexts.

LLM is increasingly experimented in the healthcare industry. For example, recent studies were conducted to see the impact of LLM in drug discovery [25], extraction of medical notes [26], prediction of diagnosis-related group [27] and diagnostics [28]. A study by Klug et al. (2024) highlighted the huge opportunities to leverage LLM in clinical decision-making systems. The advancement of LLM is parallel to the world's movements towards utilizing AI in healthcare. The Ministry of Health, Malaysia [29] for example, mentioned the use of AI and big data as essential tools for improved health care delivery in the near future. The Ministry of Health Malaysia also made digitalization, advanced data analytics and AI as the main agenda for the country's health reformation. [30]

The utilization of LLMs in QIA offers several notable advantages over traditional methodologies. Firstly, LLMs enable the automated processing and analysis of large volumes of unstructured text data, significantly reducing the time and resources required for data collection and analysis. This scalability allows healthcare institutions to extract insights from diverse sources of patient feedback in a timely and efficient manner, facilitating rapid response to emerging trends or issues.

Additionally, LLMs can identify nuanced patterns, sentiments, and themes within diverse textual data, providing deeper insights into patients' perceptions of healthcare quality and identifying areas for improvement that may have been overlooked using manual methods.

Furthermore, the integration of LLM-driven insights into quality improvement initiatives has the potential to enhance the patient-centeredness of healthcare delivery. By capturing and analyzing patients' experiences, opinions and preferences as expressed in their own words, healthcare institutions can gain a more comprehensive understanding of patient needs and priorities. This patient-centric approach enables tailored quality improvement interventions that address specific patient concerns, ultimately leading to improved patient satisfaction and outcomes.

LLM-Assisted Content Analysis (LACA)

LLM-Assisted Content Analysis (LACA) is a term coined by [31] that describes the use of Large Language Models (LLMs) to enhance and streamline qualitative content analysis. Researchers begin by inputting qualitative data—such as text from interviews or documents—into an LLM, which processes and summarizes the content to provide an initial understanding. The LLM aids in coding and categorizing the text by suggesting themes and patterns based on its advanced natural language processing capabilities. This helps in identifying and organizing key topics and underlying themes more efficiently [32].

LLMs offer several advantages in the coding process, starting with consistency and standardization. By utilizing sophisticated algorithms, they apply uniform criteria across datasets, minimizing variability and subjectivity that can arise from different human coders. This ensures coding reliability, especially when working with large datasets, where human inconsistencies can become more pronounced. Such standardization is crucial for maintaining high-quality and reliable results. [33, 34].

Another key benefit of LLMs is their scalability. These models can process and analyze vast amounts of text data much more quickly than manual methods, addressing significant resource constraints in qualitative research. The ability to handle large datasets efficiently makes LLMs an ideal solution for large-scale projects, where traditional methods may not be feasible due to

time and personnel limitations. [35]

LLMs also improve efficiency by automating the coding process, reducing the time required for annotation compared to manual coding. This can lead to substantial time savings, although the exact extent of these savings still requires further empirical investigation. The automation of this process helps streamline workflows and allows researchers to focus on other aspects of their projects. [36]

In addition to efficiency, LLMs contribute to error reduction by adhering to predefined coding rules. This minimizes inconsistencies often caused by human fatigue or differences in interpretation. While LLMs can reduce errors, careful validation against human coding is still necessary to ensure accuracy and reliability in the results. [37,38].

The potential for replicability and cost-effectiveness is another strength of LLM-based coding. When using the same models and algorithms, LLMs can produce more consistent results, making it easier to replicate studies or analyses. However, long-term reproducibility remains an area of ongoing research, as models and datasets evolve over time. [39,40]. In terms of cost-effectiveness, LLM-assisted coding can reduce the need for large human teams, particularly for large-scale projects. However, it is important to consider the initial implementation costs and the resources required for model training. Despite these costs, the long-term savings in human resources and time can make LLM-based coding a more economical solution for large research projects. [41,42].

Finally, LLMs have the capability for complex pattern detection. Trained on extensive datasets, they can identify nuanced patterns that manual coding might overlook. While this ability enhances the depth of analysis, it also requires rigorous validation to ensure that the patterns detected are meaningful and accurate. [43,44]. Each of these factors—consistency, scalability, efficiency, error reduction, replicability, cost-effectiveness, and complex pattern detection—demonstrates the potential of LLMs to significantly enhance coding processes in qualitative research.

LACA is using a large language model and, in our study, is specifically using a generative pre-trained transformer (GPT) in which prompts were transformed to trigger the model to generate responses i.e. codes / attribute labels. Other than this GPT method, Latent Dirichlet Allocation (LDA) is widely used to classify documents into topics. GPT is advantageous over LDA because GPT is context-aware and able to directly produce textual descriptions. [45]. In summary, LLM-assisted Content Analysis through LACA offers improved accuracy, efficiency, and scalability, making it a powerful tool for handling extensive and intricate content analysis tasks.

Ethical Consideration

This study involves secondary analysis of publicly available online reviews, which were anonymized and de-identified during the results visualization stage. No personal or hospital identifiers were associated with the reviews in the final results. Data processing was done only by the authorized researchers while data filtering was performed to ensure the exclusion of fake reviews using natural language processing and machine learning algorithms, which were trained on publicly available datasets such as Yelp.com. Protections were implemented during the presentation of results to minimize any potential risk of identifying individuals or revealing sensitive information. The study

methodology adheres to ethical standards concerning the use of publicly available online content for research purposes.

Flow of Research

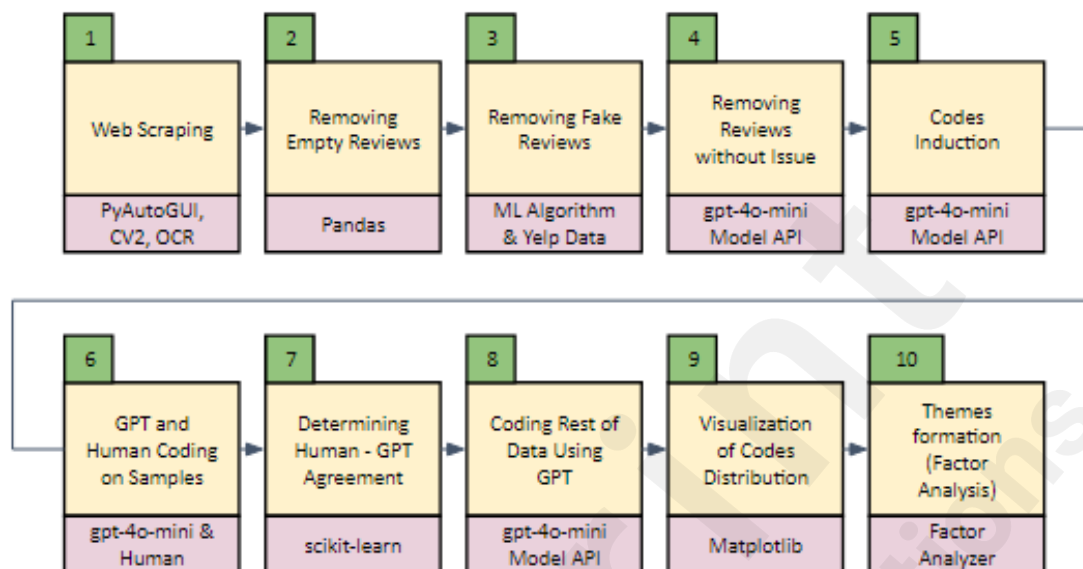


Figure 1. Flow of Research (Summarized)

1. **Web Scraping:** We begin our research by getting a list of private hospitals in the state of Selangor sourced from the Ministry of Health Malaysia's website. The list was used to search for the hospital's name on Google Search. This allowed us to locate Google Review's page for each of these hospitals. We develop a program using python PyAutoGUI module to automate data scraping. Ratings were extracted using computer vision (CV2) since star ratings in Google Review are presented in .jpg format and not in text format. Optical Character Recognition (OCR) techniques were used to detect and locate the word 'Newest' on tabs so that we could click the 'Newest' button and ensure that the data was sorted from newest to oldest. There were many reviews written too long, so part of the reviews was hidden and readers had to click on 'more' button to reveal the hidden message. We automated the process of recognizing the 'more' button by using CV2 and automatically expanding the text using PyAutoGUI click function. The process of copying all text to the computer's memory is done using PyAutoGUI select, scroll, and copy function. The copied text was stored into a .txt file with a highly specific separator between reviews.
2. **Removing Empty Reviews:** Online reviews that are not followed by comments are removed since we are doing qualitative data analysis on texts. This is done using the python Pandas module by excluding documents with empty text.
3. **Removing Fake Reviews:** Since online reviews are susceptible to manipulation by individuals from the same institution (in the case of fake positive reviews) or competitors (in the case of fake negative reviews), these fake reviews do not represent real opinions, experiences and suggestions. To do this, a program was developed based on previous studies [46, 47, 48, 49] that incorporates natural language processing methodologies with machine learning algorithms. All data for training and testing the algorithm was acquired from yelp.com, an online review platform for hotels, restaurants, hospitals among others that separate fake

reviews from real reviews. The dataset was fed into a natural language preprocessing pipeline which included the process of standardization, punctuation removal, numerical removal, tokenization, stop words removal and formation of trigrams. Each trigram is now forming a single column in a matrix of Term Frequency - Inverse Document Frequency (TF-IDF), a vector matrix that become an input for Support Vector Machine (SVM) and Logistic Regression (LR) machine learning algorithm. The ML model was trained and tested using the pre-processed yelp.com dataset and achieved precision of 0.87, recall of 0.89, high f-score and accuracy of 0.88. Using the same natural language processing pipeline, each of our hospital reviews documents is preprocessed and then transformed into vector form (TF-IDF) before being fed into our validated ML model to filter out as many fake reviews as possible.

4. **Removing Comments without Issue:** A comment can have good (positive) sentiment, bad (negative) sentiment or can have both. Our study is focusing on identifying issues (negative sentiment) related to quality in hospitals. Previous study on bias in online reviews by [50] showed that bad reviews are the most meaningful to help readers make decisions about a hospital. Other than that, we also tried to avoid having to make tuples for each label / code i.e having to label / code ('communication', 'negative') and ('communication', 'positive') instead of just 'communication' because tuples will produce double the number of variables, a phenomenon seen in article by [51]. Long list of variables also makes it more difficult during theme formation (dimension reduction) phase. We maintain reviews containing negative sentiment and filter out the remaining by calling gpt-4o-mini model API to respond 'Yes' if a review contains issues and 'No' otherwise. Full prompts in the **Appendix 1**.
5. **Codes Induction:** This stage begins with extracting issues present in each review. gpt-4o-mini model API was utilized to identify, summarize, and list issues raised by each customer based on their online customer review. Then the next step involves randomly sampling 200 of these extracts to feed the gpt-4o-mini model to produce lists of codes together with its definitions (codebook). Five iterations were performed to create a comprehensive list of codes using gpt-4o-mini model API calls. The final code list was reviewed by healthcare management experts and used as a codebook to code issues in the dataset. Full prompts in the **Appendix 2 & 3**.
6. **GPT and Human Sample Coding:** A random 200 reviews were selected. gpt-4o-mini model API were tasked to use the codebook and label the 200 reviews. A human researcher was also tasked to label the same 200 reviews. Both GPT and human coders were instructed to code each review by iterating each item in the codebook and answer 0 if the item is not an issue in the review, 1 if the item is a small issue, 2 if the item is a moderate issue, 3 if the item is a serious issue and 4 if the item is an extremely serious issue in the review (refer **Appendix 4**). If there are n_i items or codes in the codebook, the number of API calls will be $200n_i$. Inter-rater reliability was evaluated between human coder and GPT coder across the 200 reviews / documents using Cohen's Kappa, a statistical measure that accounts for chance agreement.
7. **Determining GPT - Human Agreement:** At this stage, we will have a matrix of $200 \times n_i$ where the rows are reviews / documents and the columns are item variables, i . Each cell contains integers of 0 (indicating no issue) to 4 (indicating extremely serious issue). Inter-rater reliability is a crucial aspect of ensuring the reliability of coding and categorization in our research. We do this by converting the scalar data type to binary (0 if no issue; and 1 if there is/are issue(s) related to the thematic code). Cohen's Kappa was calculated between GPT and human coder for each document. The average Kappa score across the 200 documents was calculated to assess overall agreement. Average Cohen's Kappa of more than

0.8 is acceptable to proceed to the next stage.

- 8. Coding the Rest of Data Using GPT:** As GPT coding tasks are proven to be reliable, the process of coding (described above) can then be continued solely by gpt-4o-mini model to the rest of our data. This way, we can label all our current data (and other data in future) easily without the need of human coders. This is the advantage of our current method as compared to manually coding the online reviews.
- 9. Visualization of Codes Distribution:** During this stage, we visualized the code distribution in our dataset including prevalence of each of the codes. Co-occurrences of codes are visualized through heatmaps of correlation between individual codes. Pearson Correlation was conducted to see the strength of association between thematic codes.
- 10. Themes Formation by Factor Analysis:** Thematic codes used to label review were reduced to single digit latent factors. Latent factors are essentially the underlying variables that explain the patterns of correlations among observed variables. The formation of single digit themes is important so that we could focus our effort on these themes. Although this process can be done manually through re-arrangements of codes into themes, we automate the process so that any future analysis will also be done automatically for efficiency. The use of factor analysis to reduce attributes/ codes into themes is documented by Benjamin [52]. We decided to include / exclude factors based on cumulative explained variance, Cronbach alpha [53] and qualitative assessment - content validity [54].

Overall, the methodology involves a systematic approach to collect, integrate, analyze, and interpret data from online reviews to understand the factors influencing demand for private hospitals in Selangor. Advanced techniques like machine learning and natural language processing were used to filter out fake reviews and extract meaningful insights from large datasets.

Results

Exactly 14,938 Google Review data points were scraped by our program developed using graphical user interface (GUI). These data include 53 private hospitals in the state of Selangor. The data collected consists of data from one year back from the date the data was collected (January 2023 - December 2023). Of the 14,938 data collected, a total of 12,035 (81%) evaluations were accompanied by comments while 2,903 (19%) evaluations were not accompanied by comments. Among all reviews with comments, 1,121(9.3%) reviews were fake and excluded from the data. There are 1,279 evaluations that have issues (negative sentiments) and 9,635 evaluations do not have issues (positive or neutral sentiments). The full list of thematic codes generated by LLM and its definition i.e. codebook could be accessed in **Appendix 7**. Example of actual online reviews and the codes labeled to them by LLM is shown in **Table 1** and more in **Appendix 5**.

Table 1: Example of actual online reviews and the themes assigned by LLM

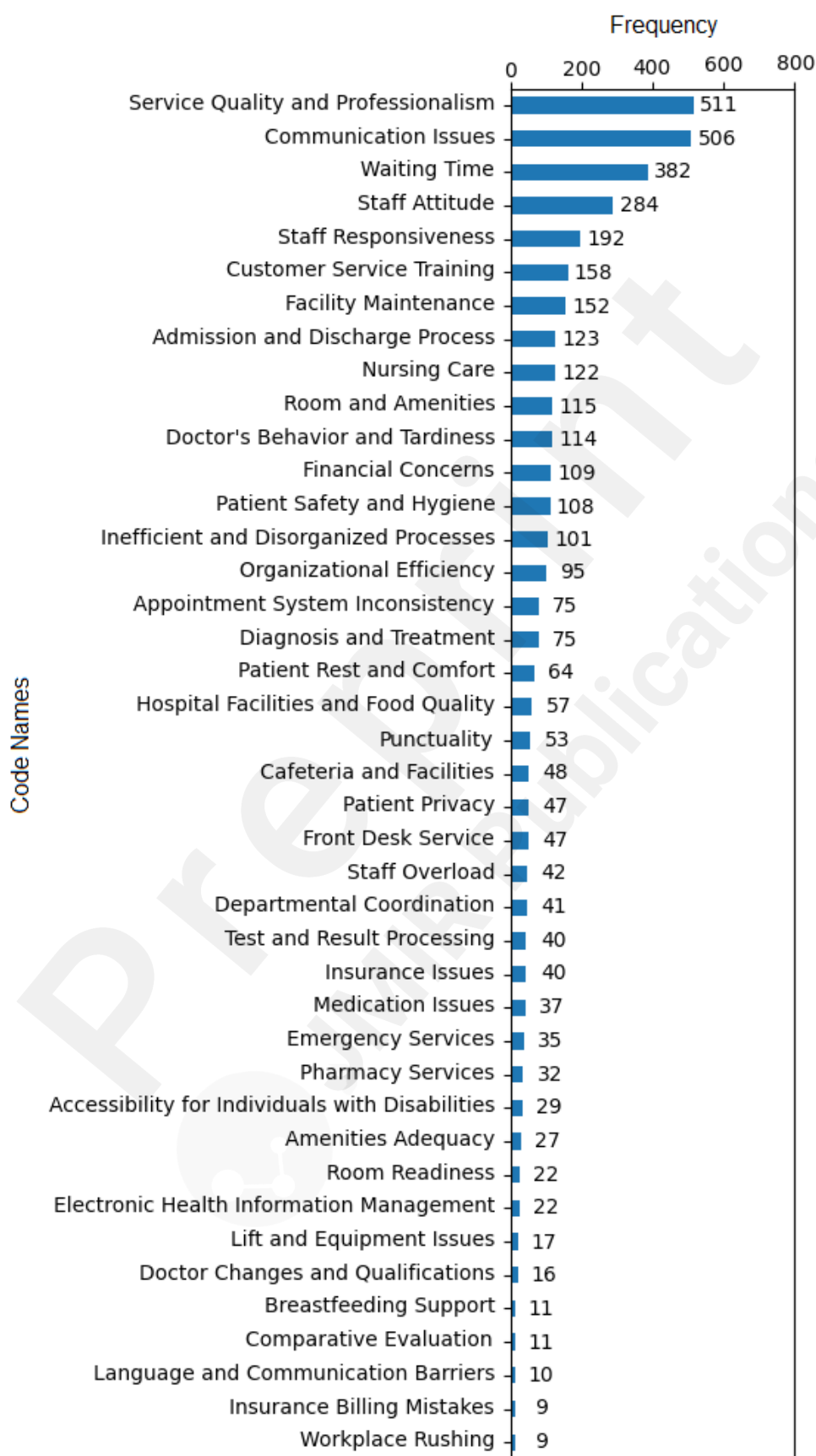
No.	Actual Online Review	Code(s) Assigned by LLM
1	Staff R*****i very helpful but <u>waiting time is too long</u> ¹ . 2 hours though. I'm patient no. 4.	1. Waiting Time

2	Very friendly staff. We were the regular there since my new born daughter always went there check up and vaccines. It's fine. But when comes to <u>serious illness, something emergency</u> ^{2,3} , they are <u>really lack of experienced staff</u> ^{1,4} .	1. Work Load 2. Emergency Services 3. Patient Safety and Hygiene 4. Doctor's Qualification and Doctor's Change
3	On 2 nd Feb 2023 this S**a nurse said will check for me for an available appointment for this specialist doctor, whom I want to see. <u>She didn't revert back to me at all</u> ^{1,2,3,4} . I called again on 28 th February this nurse A***a said my appointment was slot on 23 rd March @ 5 pm so on 22 nd March I call up to confirmed and they said my name was not in the system and I have been waited for almost 1 month. Nurse M*****h help me to rebooked but <u>didn't inform me that the date has been postponed to 24th March</u> ^{1,2} instead she told me is 3 pm so I thought is in 23 rd March at 3 pm. Such a private hospital so <u>incompetent and inefficient</u> ⁴ the nurses here. I just wants to make an appointment as this doctor specialist is always full. Simple tasks can't do it well. How you expect people will come to this hospital?	1. Communication 2. Staff Responsiveness 3. Staff Attitude 4. Inefficient and Disorganized Processes

Cohen's Kappa revealed a perfect level of agreement between the human and GPT coders. Across the 200 random reviews / documents, the lowest Kappa score between human and GPT coder was 0.44 and the highest Kappa score was 1.00. The average Cohen's Kappa score was found to be exceptionally high (0.81), indicating a very strong level of agreement beyond what would be expected by chance. These results demonstrate that the GPT coder exhibited a high degree of consistency in its coding, with minimal discrepancies across the dataset. As GPT coding tasks are validated, the process of coding was then continued solely by gpt-4o-mini model to the rest of our 1,279 data points (reviews), which gives us the following finding.

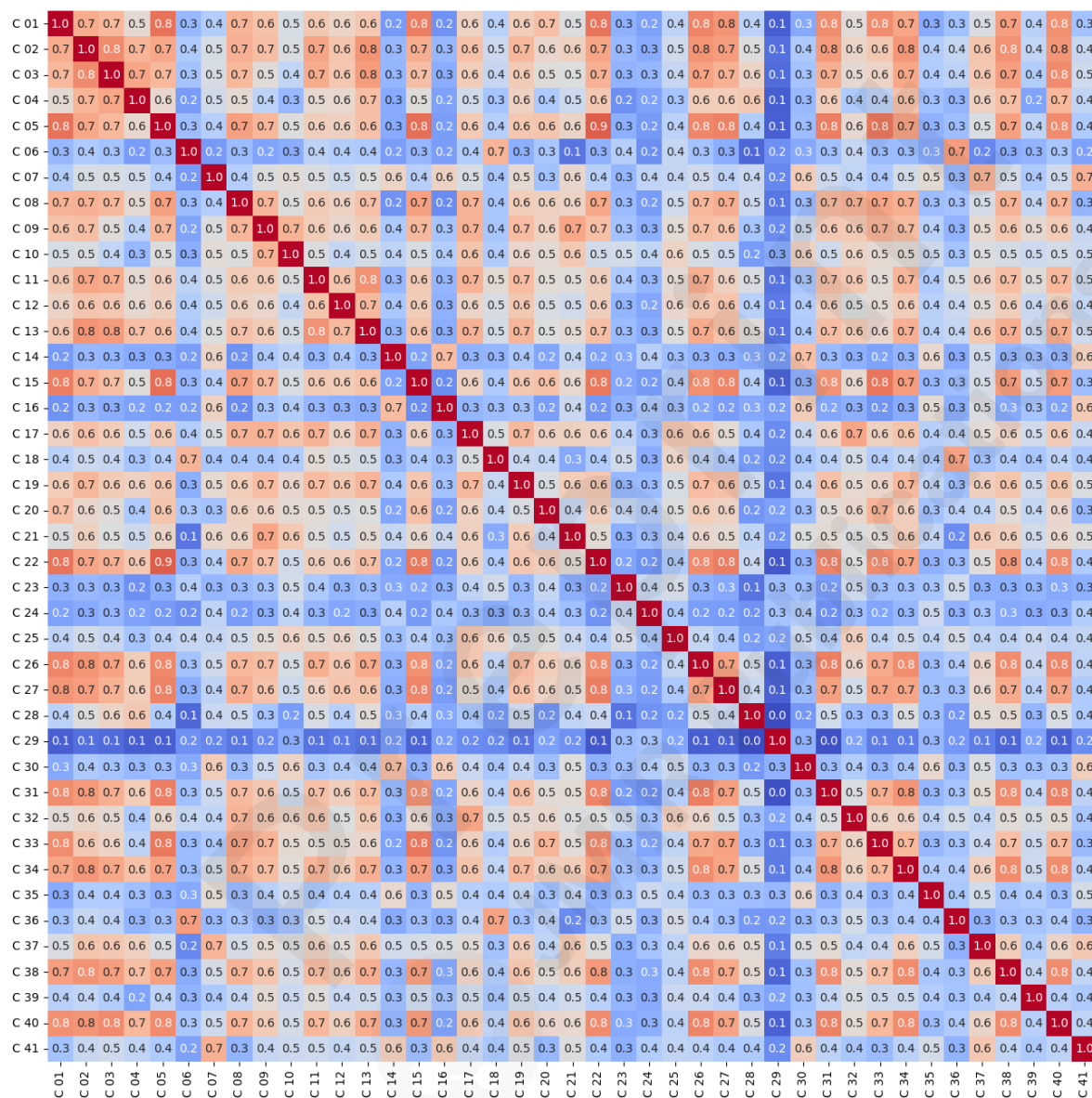
Distribution of Themes Found from Online Reviews

The most common themes based on online reviews include the themes of 'Service Quality and Professionalism' (n = 511), 'Communication' (n = 506), 'Waiting Time' (n = 382), 'Staff Attitude' (n = 284), and the theme of 'Responsiveness' (n = 192). The least common themes based on online reviews include the themes of 'Workload' (n = 9), 'Insurance Billing Errors' (n = 9), 'Communication Barriers' (n = 10), 'Comparison Between Races' (n = 11), and the theme of 'Support for Breastfeeding Mothers' (n = 11). The complete list of distributions can be seen in the distribution at **Figure 2**.

Figure 2. Distribution of thematic codes within dataset

Correlation Between Thematic Codes

The correlation was done between 2 thematic codes at a time to observe the relationship between them. We used Pearson correlation and our analysis found correlation ranges from -0.0 to 0.9 between each thematic code (**Figure 3**). The correlation between thematic codes were shown in the correlation heatmap below. After the Kaiser-Meyer-Olkin (KMO) test (0.98) and Bartlett's test of sphericity ($p < 0.05$), we accepted our data size as adequate and the code matrix as significantly different from the identity matrix and proceeded to the next step - Factor Analysis.



The heatmap shows correlation between thematic codes. As we can see, most thematic codes have a majority of red boxes indicating high correlation with most of the other thematic codes. There are however thematic codes that have the majority of blue boxes indicating low correlation with most of the other thematic codes, including these thematic codes: Financial Concerns, Facility Maintenance, Cafeteria and Facilities, Room and Amenities, Patient Privacy, Accessibility for Individuals with Disabilities, Breastfeeding Support, Lift and Equipment Issues and Insurance Billing Mistakes. The heatmap also shows some columns having similar patterns (of red and blue boxes) e.g Column 26, 27, 31 & 33, indicating that they measure an item inside the same theme and should probably be included under the same latent factor when factor analysis is done.

Factor Analysis

We were using the Factor Analyzer module from python to derive latent factors from our variables. We determined the appropriate numbers of latent factors based on eigenvalues and scree test ^[55]. Based on the total number of factors with eigenvalues more than 1, the total number of factors to include is 6. Using scree test where a straight line is drawn from the least eigenvalues to the higher eigenvalues, the suggested number of factors is also 6 (please refer **Figure 4** below). Cumulative Explained Variance was 0.74 for the first 6 factors (factors with eigenvalues more than 1). For the full list of cumulative explained variance, please refer to **Appendix 6**.

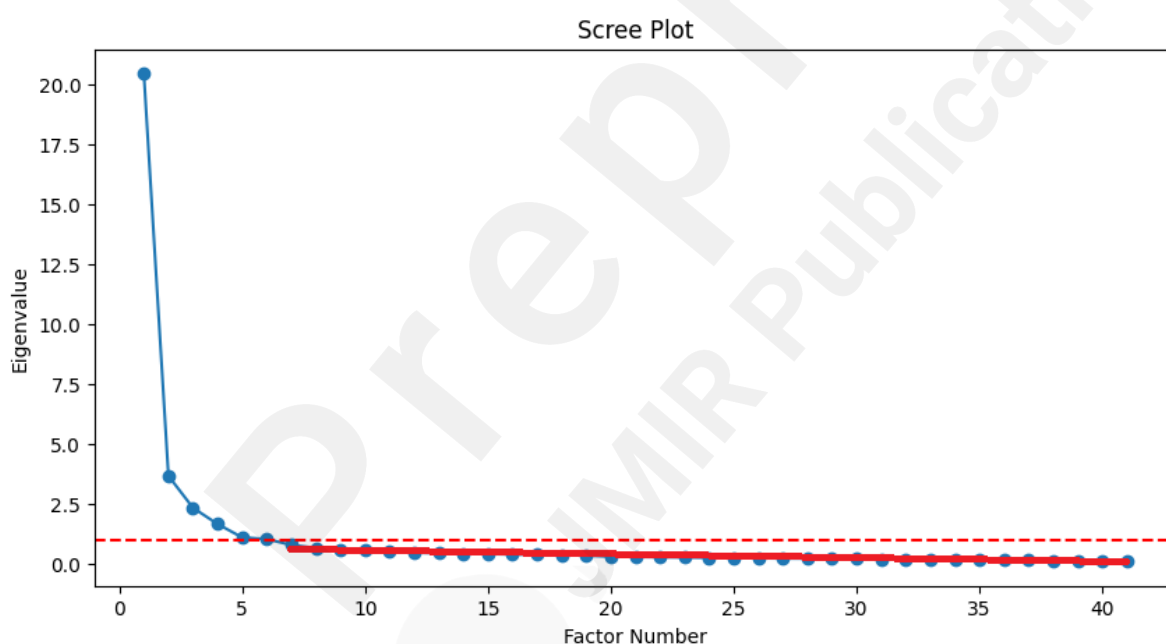


Figure 4. Scree plot of Eigenvalue against factor number.

Table 2. Latent factors together with its suggested factor labels, codes, factor loadings and Cronbach's alpha.

Code	Code Names	Factor Loading	Cronbach's Alpha
Factor 1: Service & Communication Effectiveness			
C01	Communication Issues	0.95	0.97
C05	Service Quality & Professionalism	1.00	
C08	Doctor's Behavior and Tardiness	0.51	
C15	Customer Service Training	0.97	

C20	Language and Communication Barriers	0.60	
C22	Staff Responsiveness	0.98	
C26	Departmental Coordination	0.71	
C27	Front Desk Service	0.85	
C31	Organizational Efficiency	0.80	
C33	Staff Attitude	0.98	
Factor 2: Clinical Care & Patient Experience			
C09	Nursing Care	0.77	0.92
C10	Patient Safety and Hygiene	0.64	
C11	Electronic Health Info Management	0.47	
C17	Diagnosis and Treatment	0.86	
C21	Patient Rest and Comfort	0.53	
C25	Medication Issues	0.65	
C32	Doctor Changes and Qualifications	0.72	
Factor 3: Facilities & Amenities Quality			
C07	Facility Maintenance	0.61	0.90
C14	Cafeteria and Facilities	0.89	
C16	Room and Amenities	0.87	
C30	Hospital Facilities & Food Quality	0.71	
C35	Lift and Equipment Issues	0.46	
C41	Amenities Adequacy	0.70	
Factor 4: Appointment & Patient Flow			
C03	Appointment System Inconsistency	0.42	0.89
C04	Waiting Time	0.82	
C13	Test and Result Processing	0.50	
C28	Punctuality	0.76	
Factor 5: Financial & Insurance Management			
C06	Financial Concerns	0.89	0.88
C18	Insurance Issues	0.76	
C36	Insurance Billing Mistakes	0.83	
Factor 6: Patient Rights & Accessibility			
C23	Patient Privacy	0.50	0.61
C24	Accessibility for Individuals with Disabilities	0.71	
C29	Breastfeeding Support	0.45	

Discussion

The analysis of review content identified a substantial portion of evaluations (81%) accompanied by comments, with 19% lacking comments. Of the reviews with comments, 9.3% were deemed fake and excluded, leaving 9,594 evaluations without issues and 1,279 with issues. The use of gpt-4o-mini model for coding these reviews showed a high inter-rater reliability, with an average Cohen's Kappa score of 0.81, indicating strong agreement between human and AI coders. This high level of consistency supports the validity of the coding process and the reliability of the insights derived from the data.

Granularity Issues

We acknowledged the lack of granularity of the codes produced by the gpt-4o-mini model i.e. some

of the codes are subset to the other, and some of the codes contain more than 1 topic, making it non-granular. We had noticed these results in the early phase of our research and decided to move forward with these codes. The first reason was that, in this study, we used inductive coding rather than deductive coding. Inductive coding is when codes and codebooks were created from scratch based on the available data (reviews), while deductive coding on the other hand is developing codebooks and coding the text using established concepts / theories. We did inductive coding and intentionally gave the freedom to the LLM to decide on the codes during codebook formation.

The second reason was that LLM was trained on a vast body of knowledge which is not limited to just Medicine and Management. This includes Politics, Economics, Social, Technology, Legal, Environment and Demographic domains (and this list is not exhaustive). Thus the knowledge of LLM is not restricted to one or two theories only. Due to the fact that the knowledge of LLM is cross-disciplinary, there are possibilities that we humans couldn't see the relevance of a code which is understandable only to the AI (LLM). This phenomenon is called black box problem in which mechanisms leading to an outcome are difficult to explain to humans but are effective in determining the outcome. ^[56]

Despite the granularity issues, the LLM has proven that the whole process is effective and reliable. Despite some of the codes looking non-granular to us humans, the LLM effectively uses the codes to label each review consistently across our data. For that reason, we were able to get excellent Cronbach's alpha in almost all factors during factor analysis. If the lack of granularity causes confusions to the LLM itself, the coding phase will be inconsistent or chaotic and will then result in poor factor analysis. This is proof that the producer of the codes (the LLM itself) does understand the codes precisely, despite some of them looking non-granular to us humans.

Other than that, the freedom given to LLM to decide codes without any human intervention will also make it possible for us to replicate the system in the future without human intervention using exactly the same pipeline. Thus this kind of analysis can be adapted instantaneously to any industries, and any online platforms that hold online reviews data. High replicability will make our study impacts greater.

Factor Analysis

Factor analysis identified six interpretable latent factors: 'Service & Communication Effectiveness', 'Clinical Care & Patient Experience', 'Facilities & Amenities Quality', 'Appointment & Patient Flow', 'Financial & Insurance Management', and 'Patient Rights & Accessibility'. These factors encompass the key areas influencing patient satisfaction, as reflected in the items and their factor loadings. The cumulative explained variance for the seven variables is 0.74.

As we add (or remove) factors, we face a tradeoff between cumulative explained variance, reliability measured by Cronbach's alpha and interpretability. Adding more factors increases the cumulative explained variance but compromising Cronbach's alpha on the additional factors and its interpretability. Removing factors on the other hand maintain Cronbach's alpha high on all remaining factors together with interpretability but reducing cumulative explained variance. Since any trade off required a decision maker to decide based on preference ^[57], we find it helpful to explain our preference below.

We prefer interpretable factors over higher explained variance because, unlike machine learning algorithms—e.g. Artificial Neural Networks, which can process the entire dataset to predict outcomes without fully understanding the underlying mechanisms—we humans emphasize

understanding the factors behind observed variables as a base for future studies. The need for us to understand the mechanisms behind predictions is justified by [58] who raised the concerns about black box phenomenon in machine learning algorithms, how it leads to lack of mechanism understanding and mistrust. The cumulative variance of 0.74 for our 6 factors is considered very good since our study is behavioral, as noted by [59].

Implication

The themes we identified using online review, gpt-4o-mini model, and factor analysis show strong alignment with global patterns identified in previous research. Ferreira et al.'s (2023) systematic review [60] of healthcare quality literature revealed similar priorities in patient satisfaction assessment. Their analysis showed medical care appearing in 34% of studies, communication in 31%, doctor's characteristics in 28%, accommodations in 23%, admission and discharge in 13%, nurse characteristics in 11%, appointment in 8%, environment in 8%, medical expenditure in 8%, and organization in 8%. The prominence of service quality and communication factors in our analysis mirrors these global patterns, though with some notable differences. For instance, our findings suggest a higher emphasis on appointment and waiting time issues in the Malaysian private healthcare context.

The high reliability of Service & Communication Effectiveness ($\alpha = 0.97$) coupled with strong factor loadings (0.51-1.00) demonstrates how tightly interconnected communication and service quality are in healthcare delivery. The strong correlation between communication issues and staff responsiveness ($r = 0.78$) suggests that these aspects cannot be effectively addressed in isolation. This aligns with previous findings about the integrated nature of healthcare service quality (Ferreira et al., 2023).

The emergence of distinct factors for both operational aspects (Appointment & Patient Flow, $\alpha = 0.89$) and support services (Financial & Insurance Management, $\alpha = 0.88$) indicates that patient satisfaction in private hospitals extends beyond clinical care. The moderate correlation between waiting time and staff attitude ($r = 0.41$) reveals how operational inefficiencies can impact interpersonal aspects of care delivery. The questionable reliability of the Patient Rights & Accessibility factor ($\alpha = 0.61$) points to potential measurement challenges in these areas, suggesting a need for refined assessment tools. This finding reflects the broader challenge in healthcare research of quantifying and standardizing measurements of patient rights and accessibility (Williams et al., 2010).

Limitation

Despite providing valuable insights into patient experiences, our study has several important limitations that should be considered when interpreting the results. First, the reliance on online reviews as the primary data source may introduce sampling bias, as this method potentially excludes perspectives from patients who are less likely to share feedback online, particularly older patients or those with limited digital access. While more granular categorization could have been achieved through human modulation, which might have improved the specificity and interpretability of our analysis, it also limits the true potential of LLMs, which are trained on vast bodies of knowledge. Future research could compare LACA outcomes based on codebooks entirely produced by LLMs versus those modulated by humans.

Second, the geographical scope of our study, being limited to private hospitals in Selangor, Malaysia, may not fully represent patient experiences in other regions or healthcare settings. This regional focus, while providing depth in local context, limits the generalizability of our findings to other

geographical areas or healthcare systems. Third, variations in review platforms and user demographics could affect the comprehensiveness of our findings. Different online platforms may attract distinct user demographics and encourage varying styles of feedback, potentially skewing our understanding of patient experiences [61]. Future research should address these limitations by incorporating multiple data sources, expanding geographical coverage, and considering diverse patient populations to provide a more comprehensive understanding of hospital quality issues.

Conclusion

This study underscores the effectiveness of LLM-Assisted Content Analysis for processing large-scale patient feedback, achieving high reliability (Kappa = 0.81) between human and automated content analysis. The identification of six distinct factors explaining 74% of variance provides a structured framework for understanding patient satisfaction in private hospitals.

Our findings suggest several actionable improvements for hospital management: (1) investing in integrated staff training programs focusing on communication skills and service delivery, given the strong factor loadings in these areas. (2) implementing advanced appointment systems and patient tracking technologies to address waiting time concerns, which appeared in 29.9% of negative reviews. (3), developing integrated financial service units to handle billing, insurance, and payment issues cohesively, addressing the high-severity impact of these issues when they occur.

For policy makers, our results indicate the need for more structured guidelines in patient rights and accessibility standards. Healthcare regulators might consider developing comprehensive frameworks that address these aspects more systematically. Future research should expand this analysis to other geographical regions and healthcare contexts. Additionally, investigating the long-term impact of implementing LLM-assisted content analysis on online reviews could provide valuable insights into their effectiveness for continuous quality improvement in healthcare settings.

Acknowledgements

MHS contributed in data scraping, programming, data analysis and initial manuscript writing. NM and FR contributed in supervising the project, methodology, statistical analysis and editing the paper. The project is self-funded by the researchers therefore there is no conflict of interest with any company or organization.

Data Availability

In this study, despite the data being originally derived from a public online review platform, the data used for analysis are not made publicly available to prevent the potential identification of specific hospitals. The original datasets include non-anonymized patient feedback from private hospitals in Selangor, Malaysia, which may contain sensitive information. Requests for access to the data can be made to the corresponding author, with approval subject to compliance with ethical standards and institutional review board requirements. Similar data can also be scraped from public online review platforms. The codes used in this study can be requested from the corresponding author.

Conflicts of Interest

The project is self-funded by the researchers therefore there is no conflict of interest with any company or organization.

Abbreviations

AI: Artificial Intelligence

FA: Factor Analysis

GPT: Generative Pre-trained Transformer

GUI: Graphical User Interface

HCAHPS: Hospital Consumer Assessment of Healthcare Providers and Systems

KMO: Kaiser-Meyer-Olkin (Statistical Test)

LACA: LLM-Assisted Content Analysis

LDA: Latent Dirichlet Allocation

LLM: Large Language Model

ML: Machine Learning

NLP: Natural Language Processing

OCR: Optical Character Recognition

QIA: Quality Improvement Activities

SERVQUAL: Service Quality (Questionnaire)

Appendix 1

API call for detecting presence of issue(s) in an online review.

```
def identify_issue(review):  
    # Initialize the OpenAI API client  
    client = OpenAI(  
        # This is the default and can be omitted  
        api_key= API_KEY  
    )  
    chat_completion = client.chat.completions.create(  
        messages=[  
            {  
                "role": "user",  
                "content": (  
                    f"Does the following statement contain any "  
                    f"issues:\n\nStatement:\n'{review}'\n\n"  
                    f"Answer only Yes or No"  
                )  
            },  
        ],  
        model="gpt-4o-mini",  
    )  
    r = chat_completion.choices[0].message.content.lower().strip()  
  
    return r # Result
```

Appendix 2

API call for induction of thematic codes from a list of 200 random reviews.

```
def identify_codes(sample_200):  
    # Initialize the OpenAI API client  
    client = OpenAI(  
        # This is the default and can be omitted  
        api_key= API_KEY  
    )  
    chat_completion = client.chat.completions.create(  
        messages=[  
            {  
                "role": "user",  
                "content": (  
                    f"You are a thematic analyst, "  
                    f"Now you need to produce thematic codes "  
                    f"based on the following issues:\n\n"  
                    f"Issues:\n{sample_200}\n\n"  
                    f"Please list the thematic codes "  
                    f"without explanation"  
                )  
            },  
        ],  
        model="gpt-4o" # gpt-4o for more complex work  
    )  
    r = chat_completion.choices[0].message.content.strip()  
    return r # Result
```

Appendix 3

API calls for justifying inclusion of each thematic code to develop a codebook.

```
def identify_definition(code_list, sample_200):  
    # Initialize the OpenAI API client  
    client = OpenAI(  
        # This is the default and can be omitted  
        api_key= API_KEY  
    )  
    for code in code_list:  
        chat_completion = client.chat.completions.create(  
            messages=[  
                {  
                    "role": "user",  
                    "content": (  
                        f"You are a thematic analyst. "  
                        f"Please justify why '{code}' "  
                        f"should be one of thematic code "  
                        f"in the following list of issues:\n\n"  
                        f"Issues:\n{sample_200}"  
                    )  
                },  
            ],  
            model="gpt-4o" # gpt-4o for more complex work  
        )  
        r = chat_completion.choices[0].message.content.strip()  
    return r # Result
```

Appendix 4

API call for giving a scale for each thematic codes on each online review.

```
def identify_scale(review, codebook_file):  
    list_of_codes = open_txt(codebook_file).split('\n')  
    # Initialize the OpenAI API client  
    client = OpenAI(  
        # This is the default and can be omitted  
        api_key= API_KEY  
    )  
    result_list = []  
    for code in list_of_codes:  
        chat_completion = client.chat.completions.create(  
            messages=[  
                {  
                    "role": "user",  
                    "content": (  
                        f"Given Likert scale 0 = 'Not an issue', "  
                        f"1 = 'A small issue', "  
                        f"2 = 'A moderate issue', "  
                        f"3 = 'A serious issue', "  
                        f"4 = 'An extremely serious issue', "  
                        f"is '{code}' one of the issue(s) inside "  
                        f"the following statement?\n\nStatement:\n"  
                        f"{review}\n\nAnswer the value of the "  
                        f"likert scale i.e 0 or 1 or 2 or 3 or 4"  
                    )  
                }  
            ],  
            model="gpt-4o-mini",  
        )  
        result_list.append(  
            chat_completion.choices[0].message.content.lower().strip()  
        )  
    return result_list
```

Appendix 5

Sample of actual online comments and the thematic codes assigned by LLM

No.	Actual Online Comment	Code(s) Assigned by LLM
1	Staff R****i very helpful but <u>waiting time is too long</u> ¹ . 2 hours though. I'm patient no. 4.	1. Waiting Time
2	Very friendly staff. We were the regular there since my new born daughter always went there check up and vaccines. It's fine. But when comes to <u>serious illness</u> , <u>something emergency</u> ^{2,3} , they are <u>really lack of experienced staff</u> ^{1,4} .	1. Work Load 2. Emergency Services 3. Patient Safety and Hygiene 4. Doctor's Qualification and Doctor's Change
3	On 2 nd Feb 2023 this S**a nurse said will check for me for an available appointment for this specialist doctor, whom I want to see. <u>She didn't revert back to me at all</u> ^{1,2,3,4} . I called again on 28 th February this nurse A****a said my appointment was slot on 23 rd March @ 5 pm so on 22 nd March I call up to confirmed and they said my name was not in the system and I have been waited for almost 1 month. Nurse M*****h help me to rebooked but <u>didn't inform me that the date has been postponed to 24th March</u> ^{1,2} instead she told me is 3 pm so I thought is in 23 rd March at 3 pm. Such a private hospital so <u>incompetent and inefficient</u> ⁴ the nurses here. I just wants to make an appointment as this doctor specialist is always full. Simple tasks can't do it well. How you expect people will come to this hospital?	1. Communication 2. Staff Responsiveness 3. Staff Attitude 4. Inefficient and Disorganized Processes
4	To operation director of A**C, you guys <u>really need to improve the admission process</u> ² , because I never seen the improvement of this important section after multiple times visited here: 1) <u>To do a 3 seconds Covid test, patient need to wait for 30 minutes although only 3 patients in queue</u> ¹ 2) To sign 3 pages admission document, need to wait for 25 minutes. <u>I saw only 2 staffs multitasking and very busy</u> ³ , you guys should hire more staff, not to let your client waiting and suffering in pain 3) You guys should stay overnight in the room and use all the facilities in room / bathroom of hospital that you working now, <u>you would find out what extremely lack of and what need urgently improve</u> ⁴ .	1. Waiting Time 2. Admission and Discharge Processes 3. Work Load 4. Communication
5		1. Waiting Time

	The worst and slowest hospital in the world ¹ took so long to admit patient ² , even got deposit requested to customer emergency but not act like it avoid at all cost! Doctor don't really properly check no dressing or cleaning the wounded area ⁴ , all they know only to charge customer go somewhere else, trust me counter only got one person, patient was left without attended by anyone ⁵ to ask if <u>feel uncomfortable and painful</u> have to keep asking can i have that, can you provide this and that ^{3,5}	2. Admission and Discharge Processes 3. Communication 4. Nursing Care 5. Customer Service
6	The environment is good, but the waiting time is a bit long. ¹	1. Waiting Time
7	Very bad customer service! ^{3,4} Can't even make an appointment ⁵ with health screening department. They are not responsive at all ^{1,2} .	1. Communication 2. Customer Service 3. Staff Responsiveness 4. Staff Attitude 5. Organizational Efficiency
8	Professional Doctors and nurses services is fantastic. However back end as in venepuncture, PIC (Person-In-Charge) is almost always not in her place. Pharmacy section should have more than 1 dispenser ^{1,3} . How can a premium hospital not have enough manpower ¹ . Just to wait for medications is just too long. ^{2,3} A**C is my favourite hospital but please improve on your support departments.	1. Work Load 2. Waiting Time 3. Pharmacy Services
9	Overall were good. But need to improve service ² for E&R (Probably Emergency Room - ER) department. Waiting time a bit long ¹ . Thank you.	1. Waiting Time 2. Service Quality and Professionalism
10	Very bad arrangement in clinic Dr. A***r. Appointment set at 12 pm, but still need to wait for hours ^{1,2} . Understand they are patients need emergency care, but if that so, should not place the appointment half an hour later ^{1,2,3} , should be later so that we can come later.	1. Inconsistent Appointment System 2. Waiting Time 3. Service Quality and Professionalism
11	Improve the pharmacy counter a little, the pharmacy cashier is not friendly ^{1,2} , I don't know if he is very arrogant, he is slow to take medicine ^{1,2} . Wait for more than 1 hour, because the medicine is left at the bottom, it's too late. Wait until there is no one at the pharmacy, keep on asking 5 - 6 times before looking for it, then you will get it ³ .	1. Pharmacy Services 2. Staff Attitude 3. Inefficient and Disorganized Processes
12	My mum's medical check up on 21/12. 3 results pending and supposed to be sent via email to me ¹ . Up until 27/12 still no email received until have to walk in myself ² . Excuse given - due to public holidays.	1. Test Processes and Results 2. Communication

13	Dear A**C, I was really <u>disappointed with Ms. Wong's service from the Refund Department</u>¹. My son was released from the hospital on November 6th, 2022. It's been a month, and I'm supposed to get my refund in 14 days. <u>I followed up on my refund status six times in three weeks and didn't receive any response until today</u>^{1,2}. I felt like I was begging the hospital for money. I finally called M*****e on my own and questioning why I hadn't received my refund. M*****e responded; "There are still questions from M*****e that need to be answered by the hospital. The first question was sent on November 6th, and the hospital responded on November 11th. The second query was sent out on November 11th, and there has been no response as of today. M*****e is still waiting for a response from the hospital"^{1,2}. Please expedite the process. I am formally demanding my refund back by this Friday, 9th December 2022.	1. Finance (Consultation Fees, Billing and Refund) 2. Communication
14	Phone number given has <u>nobody answer</u>^{1,2}, had been called for more than 10++ minute, operator keep telling our line was busy.	1. Communication 2. Staff Responsiveness
15	I realised now a day <u>hospital not really want entertain cash patients, hospital welcome those with medical card</u>^{1,3} which they <u>can order all sort of test and able charge without hassle</u>^{2,4}. No doubt this able help hospital generate more revenue. We believe private hospital can provide good advice in diagnosing sickness and given best solution, in reality not like this, very disappointing.	1. Finance (Consultation Fees, Billing and Refund) 2. Service Quality and Professionalism 3. Customer Service 4. Doctor's Behavior and Tardiness
16	Please <u>save paper by issuing appointment letters via email or texts</u>^{1,3}. Nurse said must have paper because of scanning purposes. If airline tickets can go paperless with clear bar codes, I'm sure Assunta Hospital can invest in better scanners. Please don't print OPA letters and save trees. We need them to tackle the CO₂ emissions. Every little helps. Thanks.	1. Organizational Efficiency 2. Facility Maintenance 3. Sustainable Practice
17	It's a shame. The hospital is good with caring and professional doctors. However your experience is marred by an <u>absolutely useless patient management system</u>^{1,2,3}. It's always down which means the <u>poor staff have to resort to manual tracking</u>². I would love to know who the vendor and IT manager are so I can avoid all places that use their system.	1. Electronic Health Information Management 2. Inter-Departmental Coordination 3. Inefficient and Disorganized Processes

18	Delivery suite, lavender, and nursery department and Dr R**u blessing for us to be delivered. No regrets came to the right place and truly grateful. We are overwhelmed. But when comes to <u>payment</u>² and <u>pharmacy</u>³ the worst service ever! Gotta <u>wait extremely long</u>¹.	1. Waiting Time 2. Finance (Consultation Fees, Billing and Refund) 3. Pharmacy Services
19	A good hospital overall. But please beware of the famous gastroenterologist there. He's the one with the most patients. Wins over his patients by being soft spoken. Gives his number to his patients but <u>doesn't give updates, does not reply or just gives one word replies</u>^{1,2}. Likes to admit patients and treats them beyond his area of speciality. <u>Doesn't update the family during the course of the admission</u>¹. Very evasive when I tried to see him after clinic hours. We went through hell when my parent was his patient. Regretted seeing him. I am sure there are better doctors around. Please dont be deceived by his demeanour.	1. Communication 2. Doctor's Behavior and Tardiness 3. Nursing Care
20	I had to <u>request many times for the bag of diapers</u>^{1,2,3} i was charged but the ward was using the packs i brought in. After many requests the missing bag of diapers appeared. Can i suggest <u>accountability for the number of diapers used per day? This will settle any potential disputes</u>⁴. And patients family are not made to feel degraded for constantly asking the already busy staff.	1. Communication 2. Service Quality and Professionalism 3. Staff Responsiveness 4. Staff Attitude

Appendix 6

Full list of cumulative explained variance from factor analysis

F01: 0.50, F02: 0.59, F03: 0.64, F04: 0.69, F05: 0.71, F06: 0.74, F07: 0.76, F08: 0.77, F09: 0.79, F10: 0.80, F11: 0.81, F12: 0.82, F13: 0.84, F14: 0.85, F15: 0.86, F16: 0.87, F17: 0.88, F18: 0.89, F19: 0.90, F20: 0.90, F21: 0.91, F22: 0.91, F23: 0.92, F24: 0.93, F25: 0.95, F26: 0.96, F27: 0.96, F28: 0.97, F29: 0.98, F30: 0.99, F31: 0.99, F32: 1.00, F33: 1.00, F34: 1.00, F35: 1.00, F36: 1.00, F37: 1.00, F38: 1.00, F39: 1.00, F40: 1.00, F41: 1.00.

Note: F = Latent Factor

Appendix 7

Codebook for guiding content analysis.

Code	Label	Definition
C01	Communication Issues	Refers to any references or instances where difficulties, barriers, or misunderstandings in communication are mentioned or observed. These issues may involve misinterpretation of messages, lack of clarity, ineffective delivery, technological problems, language barriers, or any other factors that hinder clear and successful exchange of information between individuals or groups. This code captures both verbal and non-verbal communication challenges that disrupt effective interactions.
C02	Admission and Discharge Process	Refers to any references or instances related to the procedures, protocols, or experiences involved in admitting patients to and discharging them from a healthcare setting (e.g., hospital, clinic, or care facility). This includes the administrative steps, documentation, communication, coordination among staff, patient experiences, and any challenges or improvements related to the processes of admission and discharge. It captures both formal procedures and informal aspects, such as patient satisfaction, delays, confusion, or clarity of information shared with patients and their families.
C03	Appointment System Inconsistency	Refers to any references or instances where there are irregularities, discrepancies, or inefficiencies in the scheduling or management of appointments within a system. It includes issues such as scheduling errors, double bookings, delays, miscommunication about appointment times, lack of availability, or inconsistencies in how appointments are handled across different departments or staff members. This code captures frustrations, challenges, or disruptions experienced by patients, staff, or administrators due to a lack of coordination, organization, or reliability in the appointment system.
C04	Waiting Time	Refers to any references to the amount of time

		individuals (patients, clients, or visitors) spend waiting for services, appointments, or procedures. It encompasses both the perceived and actual duration of waiting, including waiting in various contexts such as for medical consultations, treatments, or administrative processes. This code captures negative experiences related to waiting time, including frustration, dissatisfaction, perceptions of inefficiency, or understanding of delays. It also includes mentions of waiting room conditions, communication regarding wait times, or efforts to reduce waiting time.
C05	Service Quality & Professionalism	Refers to any references or instances that describe the overall quality of services provided and the level of professionalism exhibited by staff or service providers. It includes aspects such as competence, responsiveness, attention to detail, courtesy, respect, and ethical behavior in interactions with patients, clients, or customers. This code captures negative assessments of service delivery, including factors such as staff knowledge, efficiency, communication skills, empathy, and the overall effectiveness of the service. It also encompasses the organizational culture or environment that influences how services are provided and how professionals conduct themselves.
C06	Financial Concerns	Refers to any references or instances that involve concerns, challenges, or issues related to the financial aspects of services or care. It includes topics such as the cost of services, affordability, insurance coverage, payment difficulties, billing disputes, or concerns about financial accessibility. This code captures expressions of stress, anxiety, or frustration related to the financial burden of receiving care or services, as well as any discussions on how financial constraints impact decision-making or access to necessary resources or treatment.
C07	Facility Maintenance	Refers to any references or instances related to the upkeep, cleanliness, functionality, and general condition of physical spaces and infrastructure within a facility. It includes mentions of issues such as broken equipment, poor sanitation, inadequate heating or cooling, maintenance delays, and overall safety or comfort concerns within the facility. This code captures negative feedback about the condition of the

		environment, as well as any impact that facility maintenance (or lack thereof) may have on the quality of service, patient or client experiences, and staff performance.
C08	Doctor's Behavior and Tardiness	Refers to any references or instances related to the conduct, demeanor, or punctuality of doctors within a healthcare setting. It encompasses aspects such as professionalism, communication style, empathy, attentiveness, and overall attitude toward patients, as well as behaviors that may affect the quality of care. It also includes issues related to tardiness or delays in doctor appointments, such as late arrivals, extended waiting times, or lack of communication regarding delays. This code captures negative perceptions of the doctor's behavior and punctuality, and how these factors influence the patient experience.
C09	Nursing Care	Refers to any references or instances related to the care and services provided by nurses within a healthcare setting. It includes aspects such as the quality of patient interactions, the competency and professionalism of nurses, the responsiveness to patient needs, and the delivery of medical care, emotional support, and patient education. This code captures negative experiences of nursing care, highlighting areas such as attentiveness, communication, empathy, technical skills, and the overall impact of nursing care on the patient's comfort, safety, and wellbeing.
C10	Patient Safety and Hygiene	Refers to any references or instances related to the safety and cleanliness practices that directly impact the well-being of patients within a healthcare setting. It includes aspects such as infection control measures, sanitation of facilities and equipment, safe handling of medications and procedures, adherence to safety protocols, and overall hygienic conditions in patient care environments. This code captures negative feedback regarding the implementation of safety and hygiene standards, highlighting issues like cleanliness of rooms, availability of protective equipment, and the extent to which patient safety is prioritized to prevent harm or infections.
C11	Electronic Health Information	Refers to any references or instances related to the use, management, and accessibility of electronic health

	Management	records (EHR) and other digital systems for storing, retrieving, and sharing patient health information. It includes aspects such as the accuracy, security, efficiency, and user-friendliness of electronic health information systems, as well as challenges or successes in managing patient data digitally. This code captures issues related to system integration, data privacy, technical difficulties, and the effectiveness of electronic systems in supporting clinical decision-making, communication between healthcare providers, and overall patient care. It also includes the impact of technology on patient confidentiality, data sharing, and healthcare workflow.
C12	Pharmacy Services	Refers to any references or instances related to the provision and quality of pharmaceutical care and services within a healthcare setting. It includes aspects such as the dispensing of medications, accuracy in prescription fulfillment, availability of medications, patient education on drug usage, and the interaction between pharmacists and patients or healthcare providers. This code captures negative experiences related to pharmacy services, including issues such as medication errors, wait times for prescriptions, clarity of medication instructions, and the overall professionalism and effectiveness of the pharmacy team in ensuring patient safety and satisfaction with their pharmaceutical care.
C13	Test and Result Processing	Refers to any references or instances related to the procedures and experiences involved in the ordering, handling, and communication of medical tests and their results within a healthcare setting. It includes aspects such as the accuracy, timeliness, and efficiency of test processing, as well as the clarity and delivery of test results to patients or healthcare providers. This code captures challenges or successes related to delays, miscommunications, errors in results, and the overall coordination between healthcare professionals, laboratories, and patients. It also encompasses patient experiences with understanding or interpreting their test results and any follow-up actions or guidance provided based on the results.
C14	Cafeteria and Facilities	Refers to any references or instances related to the quality, availability, and conditions of food services

		(cafeteria) and other on-site facilities within a healthcare or institutional setting. It includes aspects such as the variety, nutritional quality, affordability, and accessibility of food options, as well as the overall cleanliness, comfort, and functionality of facilities such as restrooms, waiting areas, and common spaces. This code captures negative feedback regarding the environment and services provided, and how they contribute to the overall experience of patients, visitors, or staff within the facility.
C15	Customer Service Training	Refers to any references or instances related to the training programs and initiatives aimed at enhancing the customer service skills of staff within an organization. It includes aspects such as the content, quality, and effectiveness of training in areas like communication, problem-solving, empathy, conflict resolution, and professionalism when interacting with clients, patients, or customers. This code captures discussions about the impact of customer service training on staff behavior, the consistency of service delivery, and overall satisfaction of those receiving the service. It may also include perceptions of whether staff are adequately trained to handle diverse customer needs and situations.
C16	Room Amenities and	Refers to any references or instances related to the condition, comfort, and features of patient rooms or service spaces, as well as the available amenities within a healthcare or hospitality setting. It includes aspects such as room cleanliness, furniture, bed comfort, lighting, privacy, temperature control, and the availability of essential amenities (e.g., toiletries, Wi-Fi, television, phone services). This code captures negative feedback regarding the physical environment and amenities provided, and how they contribute to the overall experience, comfort, and satisfaction of patients, clients, or visitors.
C17	Diagnosis and Treatment	Refers to any references or instances related to the processes of diagnosing and treating medical conditions within a healthcare setting. It includes aspects such as the accuracy and clarity of diagnoses, the methods and approaches used in treatment, the communication of diagnosis and treatment plans to patients, and the effectiveness of prescribed treatments or interventions.

		This code captures negative experiences related to how diagnoses are made, how treatment options are explained and delivered, and the overall satisfaction or dissatisfaction with the care provided. It may also include mentions of delays, challenges in treatment effectiveness, or patient understanding of their condition and treatment plan.
C18	Insurance Issues	Refers to any references or instances related to challenges, complications, or concerns involving health insurance within a healthcare setting. It includes issues such as coverage limitations, claim denials, billing errors, pre-authorization requirements, out-of-pocket costs, or confusion regarding insurance policies. This code captures both patient and provider experiences related to navigating insurance processes, including frustrations with reimbursement procedures, communication with insurance companies, or discrepancies between services rendered and insurance coverage. It also encompasses the impact of insurance-related issues on access to care, treatment decisions, and financial stress.
C19	Emergency Services	Refers to any references or instances related to the delivery, access, and quality of emergency care within a healthcare setting. It includes aspects such as response times, the efficiency of triage systems, the availability of emergency personnel and resources, and the quality of medical treatment provided during urgent or critical situations. This code captures negative experiences with emergency services, including patient perceptions of care speed, communication with healthcare providers, the professionalism of emergency staff, and the overall effectiveness of emergency interventions in managing acute health conditions or crises.
C20	Language and Communication Barriers	Refers to any references or instances related to challenges in communication due to language differences or other barriers that affect the exchange of information between individuals. It includes issues such as misinterpretations, misunderstandings, difficulty in conveying or receiving important information, and the need for interpreters or translation services. This code captures both verbal and non-verbal communication barriers that hinder effective interaction, including the impact on service delivery, patient care, decision-

		making, and overall satisfaction. It also encompasses any strategies or efforts to overcome these challenges, such as the use of technology, bilingual staff, or cultural sensitivity training.
C21	Patient Rest and Comfort	Refers to any references or instances related to the physical and emotional comfort of patients during their stay or treatment in a healthcare setting. It includes aspects such as the quality of rest (e.g., sleep, relaxation), comfort of the patient's environment (e.g., bed, room temperature, noise levels), and the availability of amenities or support to enhance comfort (e.g., pillows, blankets, privacy). This code captures both the physical conditions that affect a patient's ability to rest and recover, as well as the emotional and psychological support provided to ensure a sense of well-being. It may also include mentions of staff responsiveness to comfort needs, as well as the overall atmosphere of the healthcare setting in promoting patient comfort.
C22	Staff Responsiveness	Refers to any references or instances related to the speed, attentiveness, and effectiveness of staff in responding to the needs, concerns, or requests of patients, clients, or visitors within a healthcare or service setting. It includes aspects such as the timeliness of responses, the willingness to assist, the quality of interaction, and the efficiency with which staff address issues or provide support. This code captures negative experiences of how staff members (e.g., healthcare providers, administrative personnel, support staff) attend to and communicate with individuals, as well as the impact of responsiveness on overall satisfaction and the quality of care or service provided.
C23	Patient Privacy	Refers to any references or instances related to the protection and respect of patient confidentiality and personal information within a healthcare setting. It includes aspects such as the safeguarding of medical records, private conversations, and any sensitive data shared between patients and healthcare providers. This code captures both the policies and practices that ensure patient privacy, as well as any concerns or breaches related to unauthorized access, disclosure of information, or lack of confidentiality. It also includes

		the ways in which healthcare staff uphold or violate privacy expectations, as well as patients' perceptions of their privacy being respected throughout their care experience.
C24	Accessibility for Individuals with Disabilities	Refers to any references or instances related to the availability and quality of accommodations and services that ensure individuals with disabilities can access and navigate healthcare or service settings. It includes aspects such as physical accessibility (e.g., wheelchair ramps, elevators, accessible restrooms), communication accessibility (e.g., sign language interpreters, assistive technologies, accessible information formats), and the provision of specialized support for patients with various disabilities (e.g., mobility, sensory, cognitive). This code captures negative experiences related to how well a healthcare or service environment accommodates individuals with disabilities, as well as any challenges, barriers, or improvements in making the environment more inclusive and accessible.
C25	Medication Issues	Refers to any references or instances related to problems or challenges encountered with medications within a healthcare setting. It includes aspects such as medication errors (e.g., incorrect dosage, wrong medication), delays in receiving medications, issues with prescription fulfillment, side effects, misunderstandings about medication instructions, and difficulties in obtaining or affording prescribed medications. This code captures both patient and healthcare provider experiences related to medication management, including concerns about the effectiveness, safety, or accessibility of medications, as well as the impact of these issues on patient health and treatment outcomes.
C26	Departmental Coordination	Refers to any references or instances related to the communication, collaboration, and integration of services across different departments within a healthcare or organizational setting. It includes aspects such as the smooth flow of information between departments, the coordination of care or services, joint decision-making processes, and the timeliness and efficiency of inter-departmental collaboration. This code captures negative experiences related to how well

		departments work together to provide comprehensive, patient-centered care, and any issues arising from lack of coordination, delays, or fragmentation of services that impact patient outcomes or operational efficiency.
C27	Front Service Desk	Refers to any references or instances related to the quality of service provided by front desk staff or receptionists in a healthcare or service setting. It includes aspects such as the professionalism, friendliness, efficiency, and helpfulness of staff in managing patient or client inquiries, appointment scheduling, check-ins, and general administrative tasks. This code captures negative experiences related to the first point of contact within a facility, including issues such as long wait times, lack of communication, or difficulties in obtaining necessary information or assistance. It also reflects how the front desk service contributes to the overall patient or client experience and satisfaction.
C28	Punctuality	Refers to any references or instances related to the timeliness of appointments, services, or staff within a healthcare or service setting. It includes aspects such as the on-time arrival of healthcare providers, timely initiation of procedures or appointments, and adherence to scheduled times. This code captures negative experiences related to punctuality, including instances of delays, waiting times, and how such issues affect patient or client satisfaction, perception of service quality, and overall experience. It also reflects the impact of punctuality on the efficiency of operations and the delivery of care or services.
C29	Breastfeeding Support	Refers to any references or instances related to the assistance, guidance, and resources provided to mothers to support breastfeeding in a healthcare or community setting. It includes aspects such as the availability of lactation consultants, education on breastfeeding techniques, emotional support, advice on managing common breastfeeding challenges (e.g., latching, supply issues, pain), and access to relevant materials or equipment (e.g., pumps, pillows). This code captures negative experiences with breastfeeding support, highlighting the effectiveness of the support offered, the professionalism and empathy of healthcare providers, and the impact on the mother's confidence, comfort,

		and success with breastfeeding.
C30	Hospital Facilities and Food Quality	Refers to any references or instances related to the physical environment of the hospital and the quality of food provided to patients and visitors. It includes aspects such as the cleanliness, comfort, and functionality of hospital facilities (e.g., patient rooms, waiting areas, restrooms, common spaces), as well as the variety, nutritional value, taste, and overall quality of food served. This code captures negative feedback about the hospital's infrastructure and food services, highlighting how these factors contribute to the overall patient experience, comfort, satisfaction, and well-being during their stay or visit.
C31	Organizational Efficiency	Refers to any references or instances related to the effectiveness and streamlined operation of an organization or healthcare setting. It includes aspects such as the smoothness of internal processes, timely delivery of services, coordination among staff, resource management, and the ability to minimize delays or disruptions. This code captures negative experiences regarding how well the organization utilizes its resources, manages workflows, and ensures that services are delivered in an organized, prompt, and cost-effective manner. It also reflects the impact of organizational efficiency on patient care, staff performance, and overall satisfaction.
C32	Doctor Changes and Qualifications	Refers to any references or instances related to changes in healthcare providers (e.g., doctor substitutions, staff turnover) and the qualifications or expertise of doctors within a healthcare setting. It includes aspects such as the introduction of new doctors, changes in the assigned physician, transitions between healthcare providers, and the qualifications, experience, or specialization of the doctors involved in patient care. This code captures negative experiences with doctor changes, including concerns about continuity of care, trust, and familiarity with the patient's medical history, as well as the perceived competence or confidence in a doctor's qualifications and their ability to provide effective care.
C33	Staff Attitude	Refers to any references or instances related to the behavior, demeanor, and general attitude of staff members within a healthcare or service setting. It

		includes aspects such as professionalism, empathy, friendliness, respect, and the overall approachability of staff, as well as their willingness to help or provide assistance. This code captures negative experiences regarding how staff interact with patients, clients, or visitors, and how their attitudes impact the overall experience. It may also encompass staff responsiveness, communication style, and the emotional tone conveyed during interactions.
C34	Staff Overload	Refers to any references or instances related to staff being overburdened or stretched beyond their capacity to effectively perform their duties within a healthcare or service setting. It includes aspects such as high patient or client caseloads, insufficient staffing levels, excessive work hours, or the pressure of managing multiple responsibilities. This code captures both the impact of staff overload on the quality of care or service, as well as any associated challenges such as burnout, delays, decreased efficiency, and staff stress. It also reflects how staff overload may affect patient or client experiences, including longer wait times, reduced attention, or diminished service quality.
C35	Lift and Equipment Issues	Refers to any references or instances related to problems or challenges with the functionality, maintenance, or availability of lifts (elevators) and other essential equipment within a healthcare or service setting. It includes aspects such as malfunctioning lifts, delays in lift access, inadequate equipment for patient handling (e.g., wheelchairs, stretchers, medical devices), and the overall reliability of tools necessary for efficient patient care and service delivery. This code captures both the impact of equipment failures on operational efficiency and patient experience, as well as the staff's ability to provide safe, timely, and effective care.
C36	Insurance Billing Mistakes	Refers to any references or instances related to errors or discrepancies in the billing process involving insurance claims. It includes issues such as incorrect charges, misapplied insurance coverage, claim denials, duplicate charges, billing for services not rendered, or delays in processing claims. This code captures both the impact of these mistakes on patients (e.g., unexpected out-of-pocket costs, confusion) and the administrative

		challenges faced by healthcare providers or insurance companies in rectifying billing issues. It also reflects patient or client frustration, the need for clarification, and any efforts to resolve billing errors.
C37	Room Readiness	Refers to any references or instances related to the preparation and condition of patient rooms or service areas prior to patient use within a healthcare or service setting. It includes aspects such as the cleanliness, availability of necessary equipment, proper functioning of room amenities (e.g., lighting, temperature control, medical devices), and overall readiness for patient care. This code captures negative experiences regarding how well rooms are prepared and maintained, as well as any delays, deficiencies, or inconveniences caused by rooms not being adequately ready for use at the time of patient arrival.
C38	Workplace Rushing	Refers to any references or instances related to the hurried or pressured environment in a workplace, where employees or service providers feel compelled to work quickly or inefficiently due to time constraints, high workloads, or organizational demands. It includes aspects such as employees rushing through tasks, interactions, or processes, potentially compromising the quality of work, attention to detail, or service delivery. This code captures both the causes of workplace rushing (e.g., understaffing, tight deadlines) and its impact on staff performance, employee stress, patient care, or customer satisfaction. It may also reflect how the rushed environment affects workplace morale and communication.
C39	Comparative Evaluation	Refers to any references or instances where individuals compare or evaluate the experiences, behaviors, or characteristics of the Chinese, Indian and Malay communities, particularly within a specific context (e.g., healthcare, education, workplace, or cultural practices). It includes aspects such as perceived differences in values, practices, communication styles, or treatment preferences between the two groups. This code captures negative evaluations, discussions, or judgments about how these communities interact with services, systems, or each other, and the implications of such comparisons on societal attitudes, cultural understanding, or service delivery. It may also reflect

		issues related to stereotypes, biases, or efforts to promote inclusivity.
C40	Inefficient and Disorganized Processes	Refers to any references or instances related to processes, systems, or workflows that are perceived as ineffective, disorganized, or lacking in structure within a healthcare or service setting. It includes aspects such as delays, confusion, poor coordination, redundant steps, lack of clear procedures, or the failure to streamline operations. This code captures both the direct impact of inefficiency and disorganization on service delivery, patient care, or customer experience, as well as staff frustrations or challenges in navigating such processes. It may also reflect suggestions for improvement or experiences of frustration due to disorganized environments that hinder productivity and satisfaction.
C41	Amenities Adequacy	Refers to any references or instances related to the sufficiency and quality of amenities provided within a healthcare or service setting. It includes aspects such as the availability, functionality, comfort, and overall appropriateness of amenities like seating, restrooms, food services, parking, Wi-Fi, and other facilities provided for patients, clients, or visitors. This code captures negative experiences regarding whether the amenities meet the needs of individuals and contribute to their overall satisfaction and comfort. It may also reflect any gaps in amenities that impact the experience of care or services provided.

References

- ¹ Lambert, M.J. and Shimokawa, K., 2016. Collecting client feedback. *Americal Psychological Association*.
- ² Gondek, D., Edbrooke-Childs, J., Fink, E., Deighton, J. and Wolpert, M., 2016. Feedback from outcome measures and treatment effectiveness, treatment efficiency, and collaborative practice: A systematic review. *Administration and Policy in Mental Health and Mental Health Services Research*, 43, pp.325-343.
- ³ Muhammad Butt, M. and Cyril de Run, E., 2010. Private healthcare quality: applying a SERVQUAL model. *International journal of health care quality assurance*, 23(7), pp.658-673.
- ⁴ Aliman, N.K. and Mohamad, W.N., 2016. Linking service quality, patients' satisfaction and behavioral intentions: an investigation on private healthcare in Malaysia. *Procedia-social and behavioral sciences*, 224, pp.141-148.
- ⁵ Abd Rashid, M., Mansor, A. and Hamzah, M., 2011. service quality and patients' satisfaction in healthcare service in Malaysia. *International Journal of Customer Service Management*, 1(1), pp.41-49.
- ⁶ Greaves, F., Lavery, A.A., Cano, D.R., Moilanen, K., Pulman, S., Darzi, A. and Millett, C., 2014. Tweets about hospital quality: a mixed methods study. *BMJ quality & safety*, 23(10), pp.838-846.
- ⁷ Hawkins, J.B., Brownstein, J.S., Tuli, G., Runels, T., Broecker, K., Nsoesie, E.O., McIver, D.J., Rozenblum, R., Wright, A., Bourgeois, F.T. and Greaves, F., 2016. Measuring patient-perceived quality of care in US hospitals using Twitter. *BMJ quality & safety*, 25(6), pp.404-413.
- ⁸ Lee, Y.U., Chung, S.H. and Park, J.Y., 2024. Online Review Analysis from a Customer Behavior Observation Perspective for Product Development. *Sustainability*, 16(9), p.3550.
- ⁹ Ranard, B.L., Werner, R.M., Antanavicius, T., Schwartz, H.A., Smith, R.J., Meisel, Z.F., Asch, D.A., Ungar, L.H. and Merchant, R.M., 2016. What can Yelp teach us about measuring hospital quality?. *Health Affairs (Project Hope)*, 35(4), p.697.
- ¹⁰ Rahim, A.I.A., Ibrahim, M.I., Musa, K.I., Chua, S.L. and Yaacob, N.M., 2021, October. Patient satisfaction and hospital quality of care evaluation in malaysia using servqual and facebook. In *Healthcare* (Vol. 9, No. 10, p. 1369). MDPI.
- ¹¹ New Straits Times, 2023. Malaysia Ranks 3rd in World for Spending Most Time Online, *nst.com.my*
- ¹² Simon Kemp (Datareportal), 2023. Digital 2023: Malaysia. *datareportal.com*
- ¹³ Lim, H.M., Wong, S.S., Yip, K.C., Chang, F.W.S., Chin, A.J.Z., Teo, C.H., Abdullah, A. and Ng, C.J., 2022. Online health information-seeking behaviour of patients attending a primary care clinic in Malaysia: a cross-sectional study. *Family Practice*, 39(1), pp.38-45.
- ¹⁴ Nor, N.A.M., Zuberi, Z.H.S., Nazari, N.S.M. and Baharuddin, N.A., 2022. Internet use, online health information seeking behavior and electronic health literacy among dental auxiliary personnel in Malaysia.
- ¹⁵ Ministry of Health Malaysia, 2023. Health White Paper for Malaysia. *Ministry of Health Malaysia*.
- ¹⁶ Braun, V. and Clarke, V., 2006. Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2), pp.77-101.
- ¹⁷ Frith, H. and Gleeson, K., 2011. Qualitative data collection: Asking the right questions. *Qualitative research methods in mental health and psychotherapy: A guide for students and practitioners*, pp.55-67.
- ¹⁸ Merriam, S.B. and Tisdell, E.J., 2015. *Qualitative research: A guide to design and implementation*. John Wiley & Sons.
- ¹⁹ Tuan, A. and Grandi, S., 2018. Emerging trends in qualitative research: a focus on Social Media. *Mercati e competitività*: 4, 2018, pp.17-26.
- ²⁰ Hassani, H., Beneki, C., Unger, S., Mazinani, M.T. and Yeganegi, M.R., 2020. Text mining in big data analytics. *Big Data and Cognitive Computing*, 4(1), p.1.
- ²¹ Stambach, D., Antoniak, M. and Ash, E., 2022. Heroes, villains, and victims, and GPT-3: Automated extraction of character roles without training data. *arXiv preprint arXiv:2205.07557*.
- ²² Vaswani, A., 2017. Attention is all you need. *Advances in Neural Information Processing Systems*.
- ²³ Xiao, Z., Yuan, X., Liao, Q.V., Abdelghani, R. and Oudeyer, P.Y., 2023, March. Supporting qualitative analysis with large language models: Combining codebook with GPT-3 for deductive coding. In *Companion proceedings of the 28th international conference on intelligent user interfaces* (pp. 75-78).
- ²⁴ Lubis, A.R., 2023. Balancing the Equation: Investigating AI Advantages, Challenges, and Ethical Considerations in the Context of GPT-3, Natural Language Processing, and Researcher Roles. *SAR Journal-Science and Research*, 6(4), pp.257-262.
- ²⁵ Oniani, D., Hilsman, J., Zang, C., Wang, J., Cai, L., Zawala, J. and Wang, Y., 2024. Emerging Opportunities of Using Large Language Language Models for Translation Between Drug Molecules and Indications. *arXiv preprint arXiv:2402.09588*.
- ²⁶ Chiang, C.C., Luo, M., Dumkrieger, G., Trivedi, S., Chen, Y.C., Chao, C.J., Schwedt, T.J., Sarker, A. and Banerjee, I., 2024. A large language model-based generative natural language processing framework fine-tuned on clinical notes

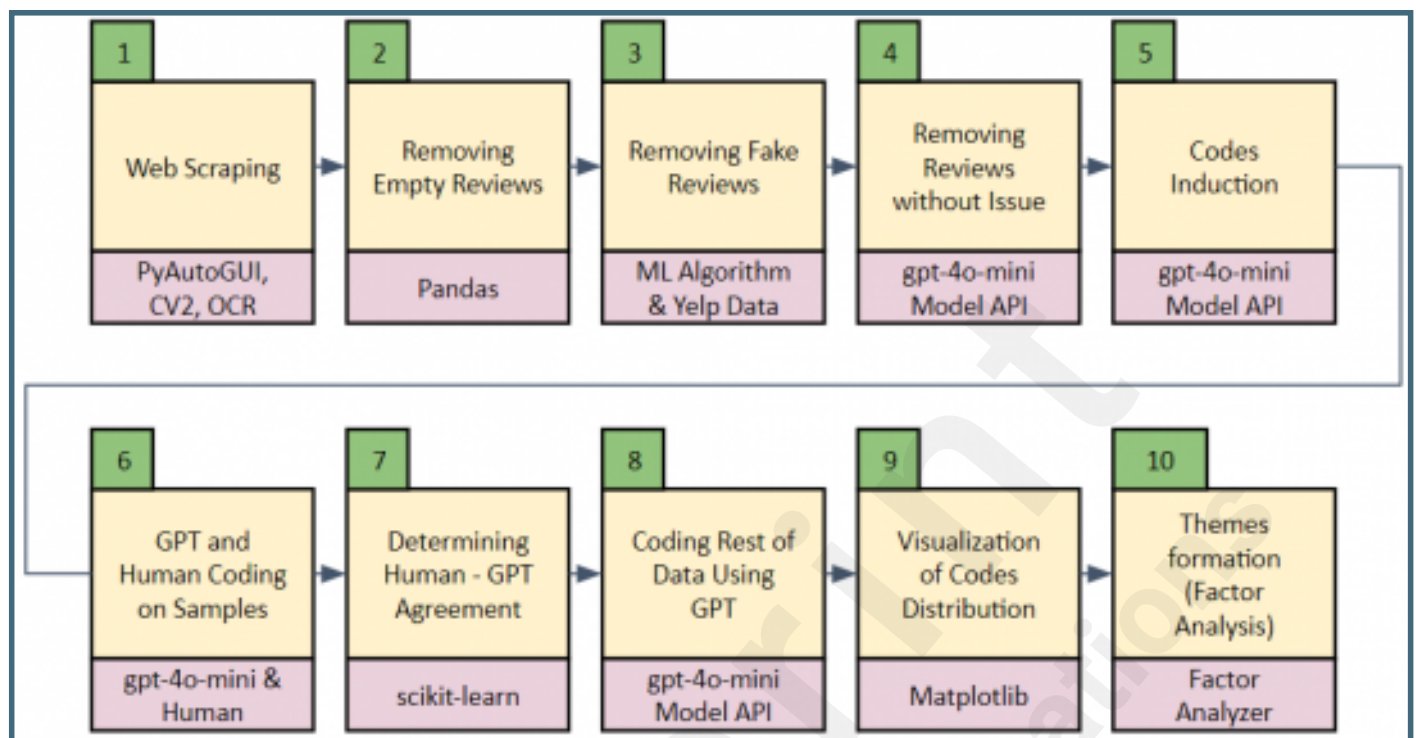
- accurately extracts headache frequency from electronic health records. *Headache: The Journal of Head and Face Pain*, 64(4), pp.400-409.
- 27 Wang, H., Gao, C., Dantona, C., Hull, B. and Sun, J., 2024. DRG-LLaMA: tuning LLaMA model to predict diagnosis-related group for hospitalized patients. *npj Digital Medicine*, 7(1), p.16.
 - 28 Gandomi, A., Wu, P., Clement, D.R., Xing, J., Aviv, R., Federbush, M., Yuan, Z., Jing, Y., Wei, G. and Hajizadeh, N., 2024. ARDSFlag: an NLP/machine learning algorithm to visualize and detect high-probability ARDS admissions independent of provider recognition and billing codes. *BMC Medical Informatics and Decision Making*, 24(1), p.195.
 - 29 World Health Organization, 2019. Regional action agenda on harnessing e-health for improved health service delivery in the Western Pacific. Chapter 4, p.22.
 - 30 Ministry of Health. 2023. Health White Paper for Malaysia. Chapter 3, p.32-55.
 - 31 Chew, R., Bollenbacher, J., Wenger, M., Speer, J. and Kim, A., 2023. LLM-assisted content analysis: Using large language models to support deductive coding. *arXiv preprint arXiv:2306.14924*.
 - 32 Khan, A.H., Kegalle, H., D'Silva, R., Watt, N., Whelan-Shamy, D., Ghahremanlou, L. and Magee, L., 2024. Automating Thematic Analysis: How LLMs Analyse Controversial Topics. *arXiv preprint arXiv:2405.06919*.
 - 33 Lewis, S.C., Zamith, R. and Hermida, A., 2013. Content analysis in an era of big data: A hybrid approach to computational and manual methods. *Journal of broadcasting & electronic media*, 57(1), pp.34-52.
 - 34 Xiao, Z., Yuan, X., Liao, Q.V., Abdelghani, R. and Oudeyer, P.Y., 2023, March. Supporting qualitative analysis with large language models: Combining codebook with GPT-3 for deductive coding. In *Companion proceedings of the 28th international conference on intelligent user interfaces* (pp. 75-78).
 - 35 Hassani, H., Beneki, C., Unger, S., Mazinani, M.T. and Yeganegi, M.R., 2020. *Text mining in big data analytics*. *Big Data and Cognitive Computing*, 4 (1), 1 [online]
 - 36 King, N. and Brooks, J., 2018. Thematic analysis in organisational research. *The SAGE handbook of qualitative business and management research methods: Methods and challenges*, pp.219-236.
 - 37 Krippendorff, K., 2018. *Content analysis: An introduction to its methodology*. Sage publications.
 - 38 McDonald, N., Schoenebeck, S. and Forte, A., 2019. Reliability and inter-rater reliability in qualitative research: Norms and guidelines for CSCW and HCI practice. *Proceedings of the ACM on human-computer interaction*, 3(CSCW), pp.1-23.
 - 39 O'Connor, C. and Joffe, H., 2020. Intercoder reliability in qualitative research: debates and practical guidelines. *International journal of qualitative methods*, 19, p.1609406919899220.
 - 40 Roberts, M.E., Stewart, B.M. and Tingley, D., 2019. Stm: An R package for structural topic models. *Journal of statistical software*, 91, pp.1-40.
 - 41 Shah, D.V., Cappella, J.N. and Neuman, W.R., 2015. Big data, digital media, and computational social science: Possibilities and perils. *The ANNALS of the American Academy of Political and Social Science*, 659(1), pp.6-13.
 - 42 Grimmer, J. and Stewart, B.M., 2013. Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political analysis*, 21(3), pp.267-297.
 - 43 Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G. and Askell, A., 2020. Language models are few-shot learners advances. In *Proceedings of the 34th International Conference on Neural Information Processing Systems* (p. 33).
 - 44 Bommasani, R., Hudson, D.A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M.S., Bohg, J., Bosselut, A., Brunskill, E. and Brynjolfsson, E., 2021. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*.
 - 45 Namita, N.R., Comparison between Traditional topic Modeling and Generative AI Topic Modeling.
 - 46 Elmogy, A.M., Tariq, U., Ammar, M. and Ibrahim, A., 2021. Fake reviews detection using supervised machine learning. *International Journal of Advanced Computer Science and Applications*, 12(1).
 - 47 Asaad, W.H., Allami, R. and Ali, Y.H., 2023. Fake Review Detection Using Machine Learning. *Revue d'Intelligence Artificielle*, 37(5).
 - 48 Alsubari, S.N., Deshmukh, S.N., Alqarni, A.A., Alsharif, N., Aldhyani, T.H., Alsaade, F.W. and Khalaf, O.I., 2022. Data analytics for the identification of fake reviews using supervised learning. *Computers, Materials & Continua*, 70(2), pp.3189-3204.
 - 49 Zhang, D., Zhou, L., Kehoe, J.L. and Kilic, I.Y., 2016. What online reviewer behaviors really matter? Effects of verbal and nonverbal behaviors on detection of fake online reviews. *Journal of Management Information Systems*, 33(2), pp.456-481.
 - 50 Roh, M. and Yang, S.B., 2021. Exploring extremity and negativity biases in online reviews: Evidence from Yelp. com.

- Social Behavior and Personality: an international journal*, 49(11), pp.1-15.
- ⁵¹ Zaman, N., Goldberg, D.M., Abrahams, A.S. and Essig, R.A., 2021. Facebook hospital reviews: Automated service quality detection and relationships with patient satisfaction. *Decision Sciences*, 52(6), pp.1403-1431.
- ⁵² Sovacool, B.K., 2013. A qualitative factor analysis of renewable energy and Sustainable Energy for All (SE4ALL) in the Asia-Pacific. *Energy Policy*, 59, pp.393-403.
- ⁵³ George, D., & Mallery, P. (2003). *SPSS for Windows Step by Step: A Simple Guide and Reference*. 11.0 Update (4th ed.). Boston: Allyn & Bacon.
- ⁵⁴ Yusoff, M.S.B., 2019. ABC of content validation and content validity index calculation. *Education in medicine journal*, 11(2), pp.49-54.
- ⁵⁵ Williams, B., Onsman, A. and Brown, T., 2010. Exploratory factor analysis: A five-step guide for novices. *Australasian journal of paramedicine*, 8, pp.1-13.
- ⁵⁶ Wang, Y., 2023. Deciphering the enigma: A deep dive into understanding and interpreting llm outputs. *Authorea Preprints*.
- ⁵⁷ Knetsch, J.L., 1989. The endowment effect and evidence of nonreversible indifference curves. *The american Economic review*, 79(5), pp.1277-1284.
- ⁵⁸ Vamathevan, J., Clark, D., Czodrowski, P., Dunham, I., Ferran, E., Lee, G., Li, B., Madabhushi, A., Shah, P., Spitzer, M. and Zhao, S., 2019. Applications of machine learning in drug discovery and development. *Nature reviews Drug discovery*, 18(6), pp.463-477.
- ⁵⁹ Williams, B., Onsman, A. and Brown, T., 2010. Exploratory factor analysis: A five-step guide for novices. *Australasian journal of paramedicine*, 8, pp.1-13.
- ⁶⁰ Ferreira, D.C., Vieira, I., Pedro, M.L., Caldas, P. and Varela, M., 2023, February. Patient satisfaction with healthcare services and the techniques used for its assessment: a systematic literature review and a bibliometric analysis. In *Healthcare* (Vol. 11, No. 5, p. 639). MDPI.
- ⁶¹ Ranard, B.L., Werner, R.M., Antanavicius, T., Schwartz, H.A., Smith, R.J., Meisel, Z.F., Asch, D.A., Ungar, L.H. and Merchant, R.M., 2016. What can Yelp teach us about measuring hospital quality?. *Health Affairs (Project Hope)*, 35(4), p.697.

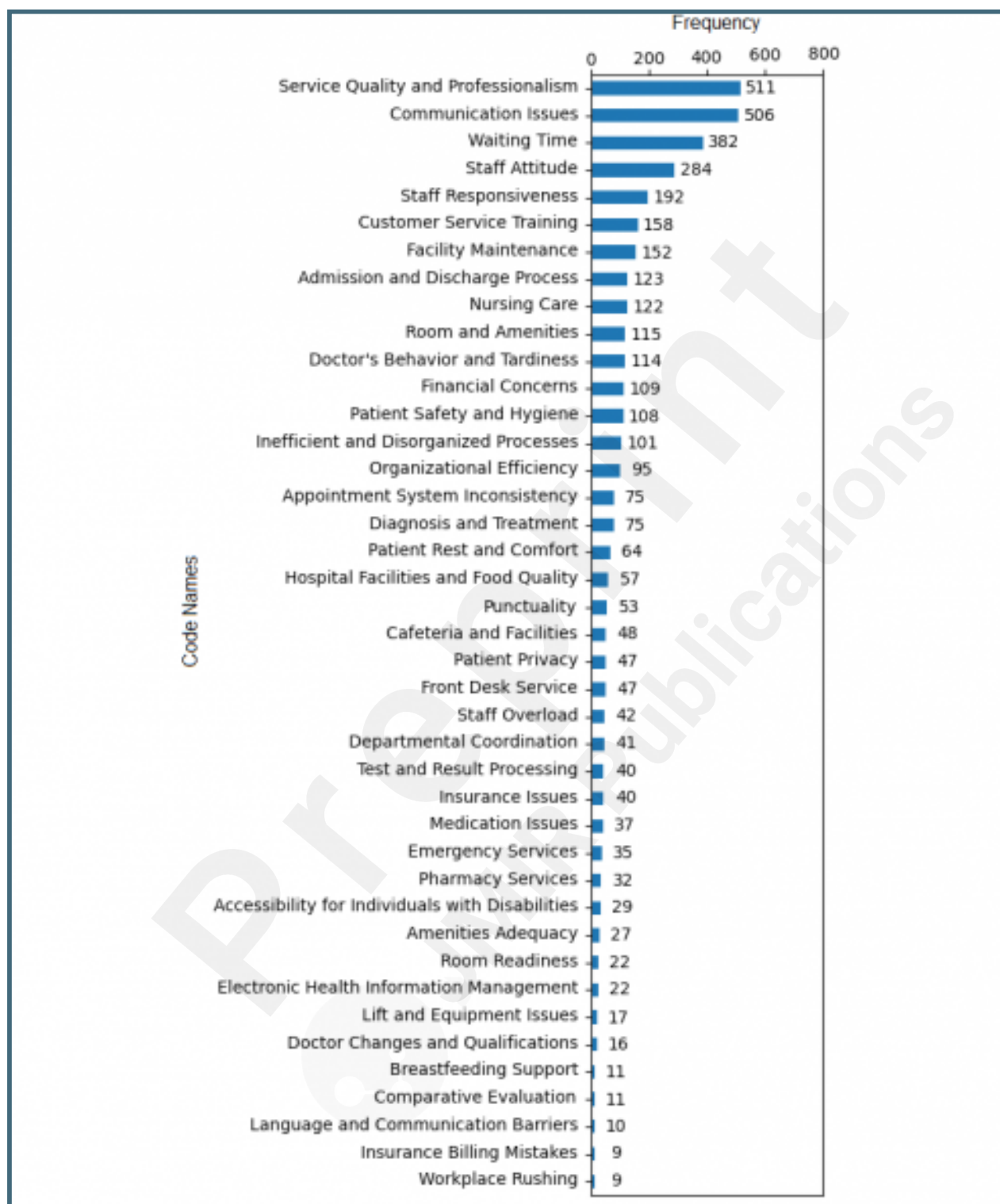
Supplementary Files

Figures

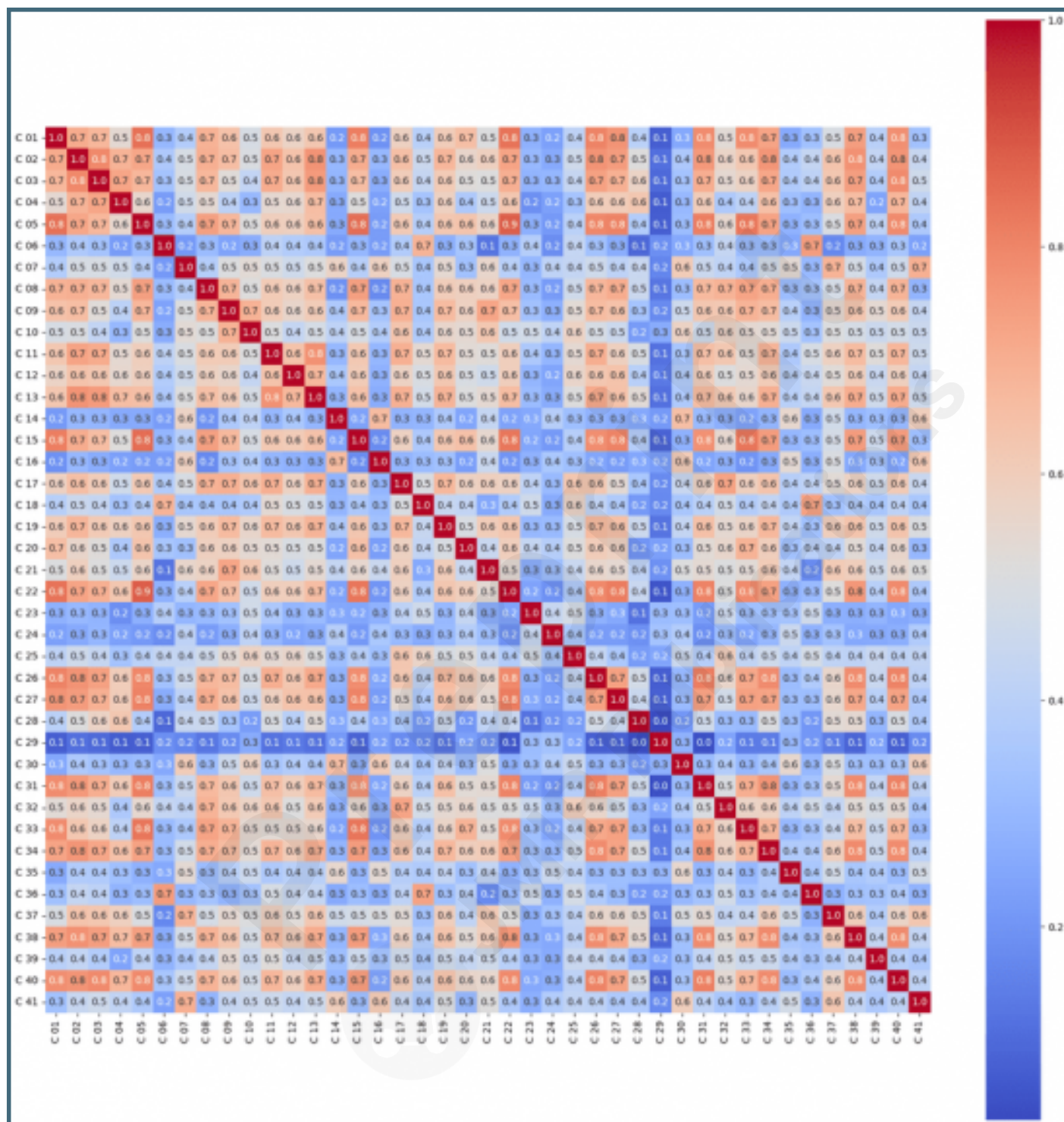
Untitled.



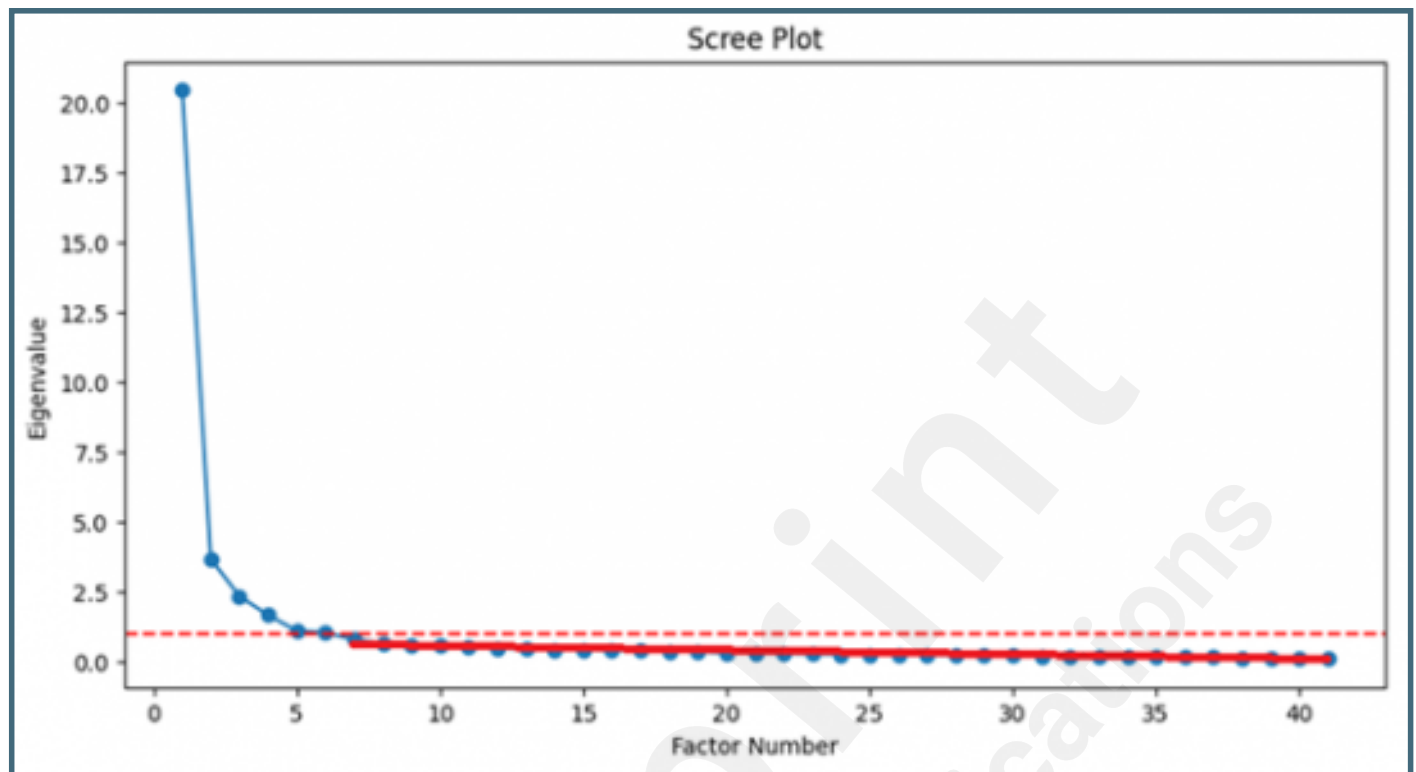
Untitled.



Heatmap of correlations between thematic codes.



Scree plot of Eigenvalue against factor number.



Multimedia Appendixes

API call for detecting presence of issue(s) in an online review.

URL: <http://asset.jmir.pub/assets/56f6ddd029dd711986079b80ce48fe20.docx>

API call for induction of thematic codes from a list of 200 random reviews.

URL: <http://asset.jmir.pub/assets/4f433bd1eb29252f67142eea0a3b2ec8.docx>

API calls for justifying inclusion of each thematic code to develop a codebook.

URL: <http://asset.jmir.pub/assets/87173ec9a657e6edcfdeea55572ef965.docx>

API call for giving a scale for each thematic codes on each online review.

URL: <http://asset.jmir.pub/assets/a4957d5bbe5e7be8aee67aad8cfdbda.docx>

Sample of actual online comments and the thematic codes assigned by LLM.

URL: <http://asset.jmir.pub/assets/64c50fbde81ae3f41a8055fd69d2934b.docx>

Full list of cumulative explained variance from factor analysis.

URL: <http://asset.jmir.pub/assets/30091f1045c8c192c96324a75b642bfa.docx>

Codebook for guiding content analysis.

URL: <http://asset.jmir.pub/assets/1dfa38d130912cb290d191ded4d45137.docx>