LLM-Assisted Content Analysis (LACA) on Online Reviews for Hospital Quality Improvements Activities

Muhammad Hafiz Sulaiman, Nora Muda, Fatimah Abdul Razak

Submitted to: Journal of Medical Internet Research
on: November 21, 2024

Table of Contents

Original Manuscript.. 5

LLM-Assisted Content Analysis (LACA) on Online Reviews for Hospital Quality Improvements Activities

Muhammad Hafiz Sulaiman[1, 2] MD, MS; Nora Muda[1] PhD; Fatimah Abdul Razak[1] PhD

[1]Department of Mathematical Sciences National University of Malaysia Bangi MY
[2]Sultan Idris Shah Hospital Serdang MY

Corresponding Author:
Muhammad Hafiz Sulaiman MD, MS
Department of Mathematical Sciences
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 article is intended to explore and set a recommendation on how LACA can be applied on online reviews to address quality issues in hospitals.

Methods: Online reviews for 53 private hospitals in Selangor Malaysia were acquired. Fake reviews were filtered out and a sample of 200 reviews was randomly extracted and fed into gpt-4o-mini model API to produce a codebook which was then used to code (label) all reviews in the dataset. Patterns of thematic codes across the whole dataset were presented using python matplotlib. The thematic codes were then summarized into themes using factor analysis to increase interpretability.

Results: 14,938 online reviews were acquired in which 1,279 comments were detected to mention quality issues. 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 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 & processing pipeline made of python selenium, gpt-4o-mini model API, and factor analyzer module can run valid and reliable thematic analysis. Despite online reviews being subject to collection and information bias, insights from online analysis is real-time and can assist hospital managers to develop hypotheses and react to trending quality issues quickly once data collection & processing pipeline is established

(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 v

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Original Paper

Dr. Muhammad Hafiz Sulaiman, MD, MS[1,2]
Associate Prof. Dr. Nora Muda, PhD[1]
Associate Prof. Dr. Fatimah Razak, PhD[1]

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

LLM-Assisted Content Analysis (LACA) on Online Reviews for Hospital Quality Improvements Activities.

Abstract

Introduction: 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. This article is intended to explore and set a recommendation on how LACA can be applied on online reviews to address quality issues in hospitals. **Method:** Online reviews for 53 private hospitals in Selangor Malaysia were acquired. Fake reviews were filtered out and a sample of 200 reviews was randomly extracted and fed into gpt-4o-mini model API to produce a codebook which was then used to code (label) all reviews in the dataset. Patterns of thematic codes across the whole dataset were presented using python matplotlib. The thematic codes were then summarized into themes using factor analysis to increase interpretability. **Results:** 14,938 online reviews were acquired in which 1,279 comments were detected to mention quality issues. 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 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 & processing pipeline made of python selenium, gpt-4o-mini model API, and factor analyzer module can run valid and reliable thematic analysis. Despite online reviews being subject to collection and information bias, insights from online analysis is real-time and can assist hospital managers to develop hypotheses and react to trending quality issues quickly once data collection & processing pipeline is established.

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

Introduction

Quality Improvement Activities in Hospitals

Getting feedbacks from patients and families is important to continuously improve patient care and ensuring patients and families satisfaction within healthcare settings (Lambert et al. 2011[1] & Godek et al 2016[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 Muhammad et al. (2010)[3], Aliman et al. (2016)[4], and Abd Rashid et al. (2004)[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 (Greaves 2014[6], Hawkins 2016[7]). Since these methods require a huge amount of time, the number of respondents for interviews or time of observation is limited thus leading to biased conclusions. 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. (Yeong Un Lee 2024[8])

Use of online reviews to get feedback on patients' experiences and opinions can assist hospital managers tackle the above limitations. Unlike ethnographic studies and interviews, online reviews by patients and families have no spatial limitations which means that patients and families can post their feelings, experiences and opinions throughout their journey before, during and after hospital stay/visit. Online reviews are also readily available on the internet and with correct tools, the analysis will involve as many respondents as possible therefore reducing bias from limited samples.

Benjamin L Ranard 2016[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.

Afiq Izzudin (2021)[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. Afiq 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 resulted in an unprecedented proliferation of publicly available data including online reviews. By 2020, global data usage was expected to surge 44 times, reaching 35.2ZB, highlighting the exponential growth of data on the internet. (Cai et al. 2018)[11] Online data include those reflecting patients' experiences, opinions, and perceptions regarding healthcare facilities. These data encompass a diverse array of sources, including social media platforms, online review websites, healthcare forums, and patient blogs.

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 alpha of > 0.70 on all factors with interpretable items, **(3)** The themes of current issues expected to be identified in private hospitals in Selangor, Malaysia follow global pattern of issues in hospital identified from literature.

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 on Braun & Clarke (2006)[12]

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. (Freeth and Gleeson 2004)[13]

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. (Meriam et al. 2015)[14] 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 (Tuan A. et al. 2018)[15]

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. Hassani H. et al. (2020)[16] 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 (Stammbach et al. 2022)[17], 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. (Vaswani 2017[18]) Leveraging advanced natural language processing (NLP) techniques, LLMs excel in tasks such as text summarization, thematic analysis (Xiao et al. 2023)[19] and sentiment analysis (Lubis A.R 2023)[20], making them well-suited for analyzing qualitative data in healthcare contexts.

LLM is increasingly experimented in the healthcare industry. For example, recent studies was conducted to see the impact of LLM in drug discovery (Oniani 2024[21]), extraction of medical notes (Chiang et al. 2024[22]), prediction of diagnosis-related group (Wang et al. 2024)[23] and diagnostics (Gandomi et al. 2024)[24]. 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. World Health Organization (2019[25]) 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. (MOH, 2023[26])

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 Robert Chew (2023)[27] 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 (Awais Hameed Khan 2024)[28].

LACA has advantages over traditional manual coding. LACA leverages advanced language models to significantly enhance content analysis, providing several notable benefits:

1. Ensures consistency and standardization by using sophisticated algorithms that apply uniform criteria across datasets, thereby minimizing the variability and subjectivity that can arise from different human coders.

2. Excels in scalability, processing and analyzing large volumes of text data quickly, which is a significant improvement over manual methods that are often constrained by time and human resources.

3. Enhances speed and efficiency by automating the coding process, substantially reducing the time required for annotation compared to the slower, more labor-intensive manual coding.

4. Reduces potential errors by strictly following predefined coding rules (codebook), avoiding the inconsistencies and mistakes that can occur with human coders who may be affected by fatigue or varying interpretations.

5. Results from LACA are consistent and replicable, provided the same models and algorithms are used, ensuring reliable outcomes across different analyses.

6. More cost-effective in the long run, especially for large-scale projects, as it reduces the need

for extensive human resources and minimizes overall time investment.

7. LLM is exposed to big data during its training phase. With a broad perspective, LACA can detect and analyze complex patterns and nuances in text that manual coding might miss.

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. (Namita 2023[29]). 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.

Flow of Research

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 were 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 Reviews Not Followed by Comments:** Online reviews that are not followed by comments are removed since we are doing qualitative data analysis on texts. This is easily done using python built-in functions 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 (Ahmed M Elmogy 2021[30], Wessam Hameed 2023[31], Saleh Nagi 2021[32] and Dongsong Zhang 2016[33]) 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 with Positive Sentiment**: A comment can have good (positive) sentiment or bad (negative) sentiment. Our study is focusing on identifying issues (negative sentiment) related to quality in hospitals. Previous study on bias in online reviews by Minjung Roh et al. 2021[34] 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 amount of variables, a phenomena seen in article by Nohel Zaman et al. 2020[35]. Long list of variables also makes it more difficult during theme formation (dimension reduction) phase. We maintain negative sentiment and filter out positive ones 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 x 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. **Codes Visualization**: 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 K Sovacool (2013)[36]. We decided to include / exclude factors based on cumulative explained variance, Cronbach alpha (George & Mallery, 2003)[37] and qualitative assessment - content validity (Yusof M.S.B 2019)[38].

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).

We randomly sampled 200 of these 1,279 reviews 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. The full list of code is shown in **Figure 1**.

A random 200 reviews were selected. gpt-4o-mini model API were tasked to use the codebook and label the 200 reviews using scale of 0 (if the thematic code is not an issue) - 4 (if the thematic code is an extremely serious issue). The scale was then converted to binary 0 (if the thematic code is not an issue) and 1 (if the thematic code is an issue regardless of the magnitude). A human researcher was also tasked to label the same 200 reviews binarily. 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. 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

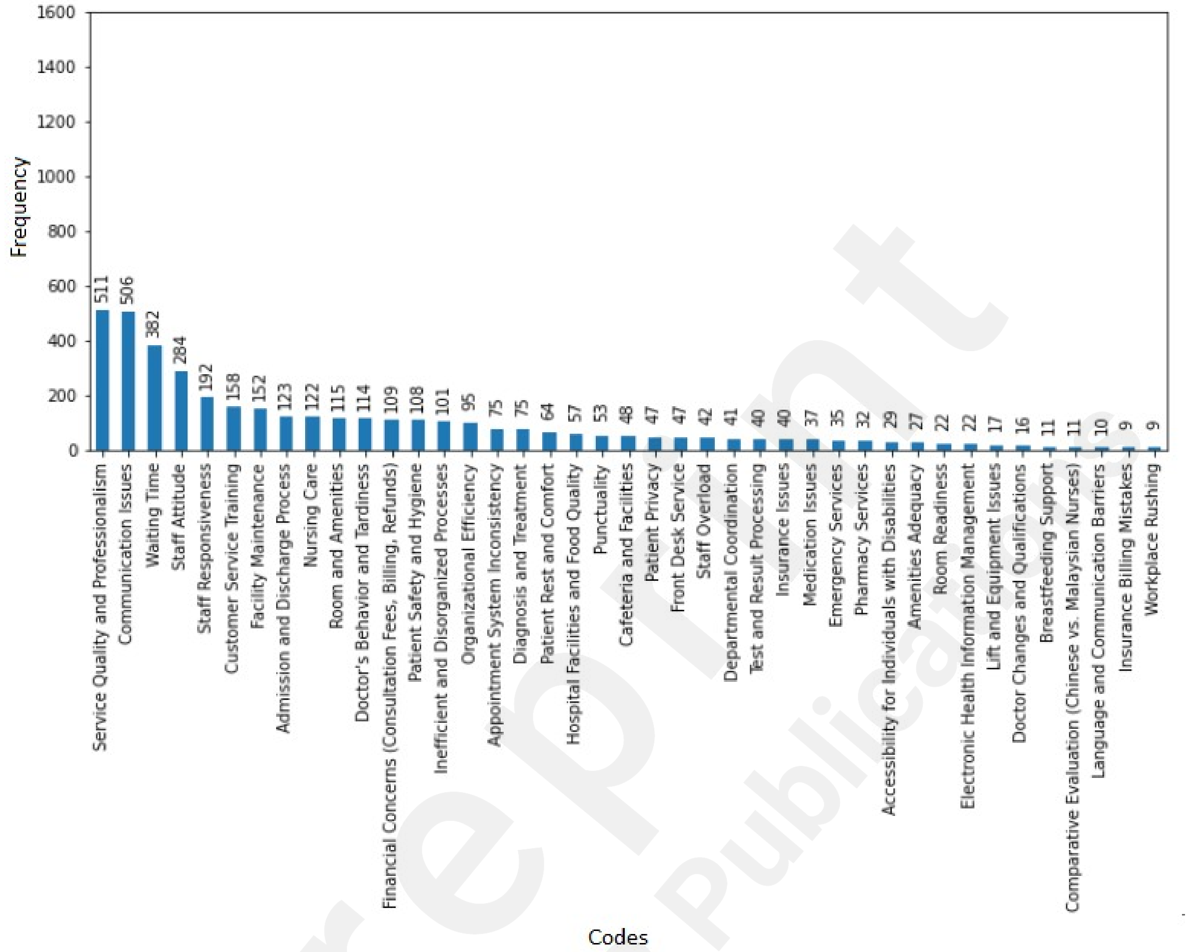
No.	Actual Online Review	Code(s) Assigned by LLM
1	Staff R****i very helpful but waiting time is too long[1]. 2 hours though. I'm patient no. 4.	1. Waiting Time
2	Very friendly staff. We were the regular there since my new born daughter always went there check up and vaccines. It's fine. But when comes to serious illness, something emergency[2,3], they are really lack of experienced staff[1,4].	1. Work Load 2. Emergency Services 3. Patient Safety and Hygiene 4. Doctor's Qualification and Doctor's Change
3	On 2nd Feb 2023 this S**a nurse said will check for me for an available appointment for this specialist doctor, whom I want to see. She didn't revert back to me at all[1,2,3,4]. I called again on 28th February this nurse A***a said my appointment was slot on 23rd March @ 5 pm so on 22nd 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 didn't inform me that the date has been postponed to 24th March[1,2] instead she told me is 3 pm so I thought is in 23rd March at 3 pm. Such a private hospital so incompetent and inefficient[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 1**.

Figure 1. 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 2**). 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.

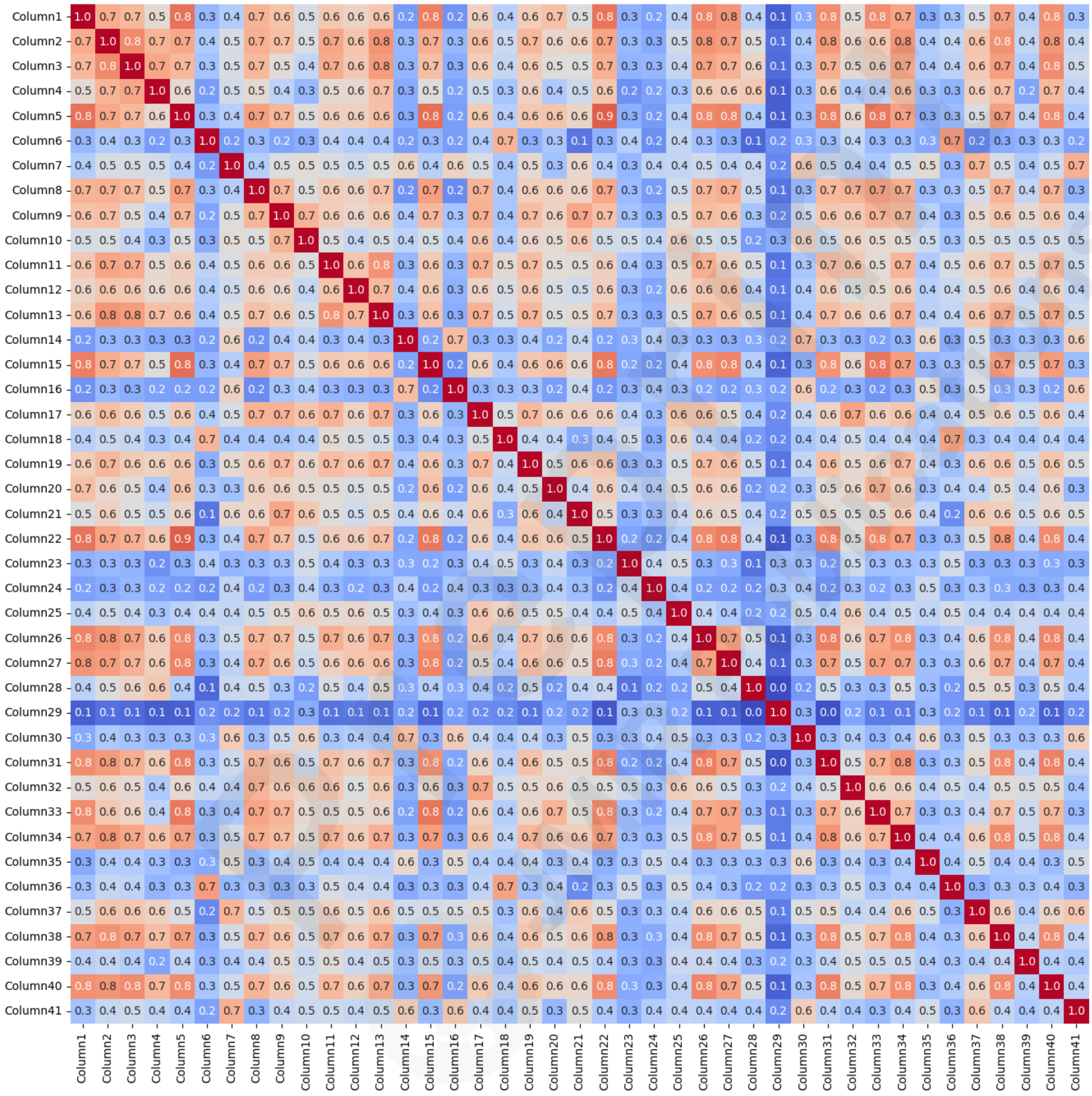


Figure 2. Heatmap of correlations between thematic codes

Column name: Column1 - Communication Issues; Column2 - Admission and Discharge Process; Column3 - Appointment System Inconsistency; Column4 - Waiting Time; Column5 - Service Quality and Professionalism; Column6 - Financial Concerns (Consultation Fees, Billing, Refunds); Column7 - Facility Maintenance; Column8 - Doctor's Behavior and Tardiness; Column9 - Nursing Care; Column10 - Patient Safety and Hygiene; Column11 - Electronic Health Information Management; Column12 - Pharmacy Services; Column13 - Test and Result Processing; Column14 - Cafeteria and Facilities; Column15 - Customer Service Training; Column16 - Room and Amenities; Column17 - Diagnosis and Treatment; Column18 - Insurance Issues; Column19 - Emergency Services; Column20 - Language and Communication Barriers; Column21- Patient Rest and Comfort; Column22 - Staff Responsiveness; Column23 - Patient Privacy; Column24 - Accessibility for Individuals with Disabilities; Column25 - Medication Issues; Column26 - Departmental Coordination; Column27 - Front Desk Service; Column28 - Punctuality; Column29 - Breastfeeding Support; Column30 - Hospital Facilities and Food Quality; Column31 - Organizational Efficiency; Column32 - Doctor Changes and Qualifications; Column33 - Staff Attitude; Column34 - Staff Overload; Column35 - Lift and Equipment Issues; Column36 - Insurance Billing Mistakes; Column37 - Room Readiness; Column38 - Workplace Rushing; Column39 - Comparative Evaluation (Chinese vs. Malaysian Nurses); Column40 - Inefficient and

Disorganized Processes; Column41 - Amenities Adequacy.

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 (Williams et al. 2010)[39]. 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 3** below).

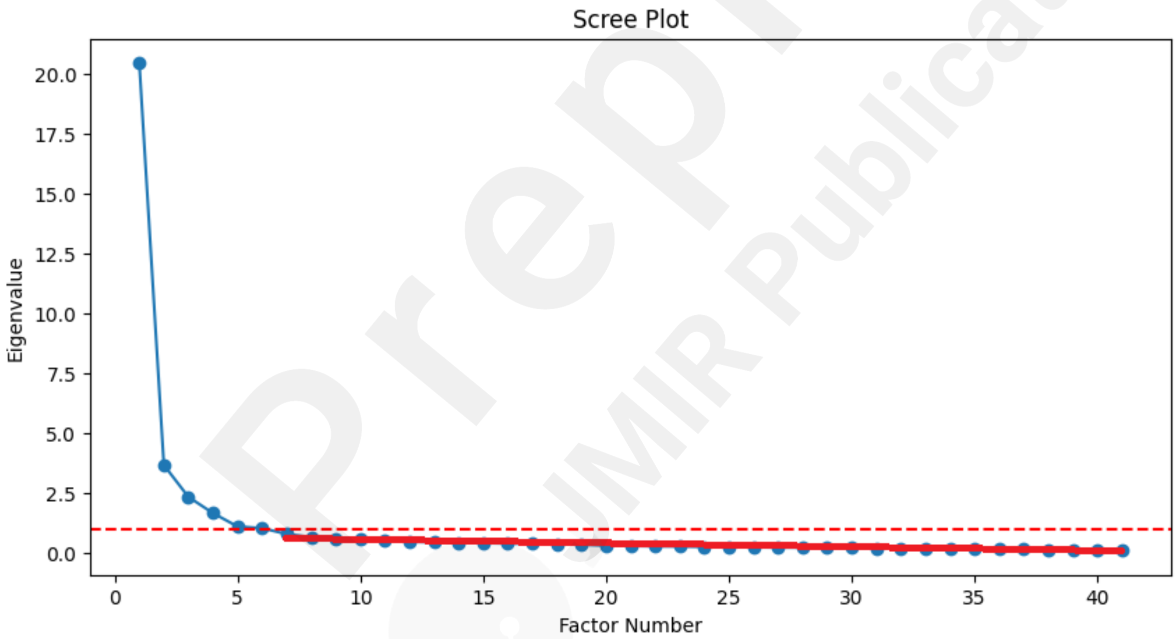


Figure 3. Scree plot of Eigenvalue against factor number.

Cumulative Explained Variance was 0.74 for the first 6 factors (factors with eigenvalues more than 1). Full list of cumulative explained variance is stated as F1: 0.50, F2: 0.59, F3: 0.64, F4: 0.69, F5: 0.71, F6: 0.74, F7: 0.76, F8: 0.77, F9: 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.

Table 2. Latent factors (F1 - F8) identification and factor loadings. The factor loadings for variables responsible to label each factor is underlined.

Code	Label	Factor Loadings					
		F1	F2	F3	F4	F5	F6
C01	Communication Issues	**0.95**	-0.07	-0.05	-0.01	0.03	0.03
C02	Admission and Discharge Process	0.42	0.13	-0.02	**0.42**	0.07	-0.03
C03	Appointment System Inconsistency	0.36	-0.08	-0.07	0.63	0.07	0.10
C04	Waiting Time	0.24	-0.23	-0.01	**0.82**	-0.02	0.08
C05	Service Quality & Professionalism	**1.00**	-0.03	0.05	-0.13	0.02	-0.05
C06	Financial Concerns	0.10	-0.05	0.08	-0.17	**0.89**	-0.05
C07	Facility Maintenance	-0.04	0.12	**0.61**	0.26	-0.04	0.00
C08	Doctor's Behavior and Tardiness	**0.51**	0.44	-0.22	0.12	-0.08	0.04
C09	Nursing Care	0.39	**0.77**	0.13	-0.23	-0.18	-0.11
C10	Patient Safety and Hygiene	0.16	**0.64**	0.19	-0.22	-0.13	0.18
C11	Electronic Health Info. Managemt.	0.07	**0.47**	-0.07	0.37	0.18	-0.08
C12	Pharmacy Services	0.26	0.24	0.09	0.20	0.21	-0.10
C13	Test and Result Processing	0.11	0.39	-0.09	**0.50**	0.10	-0.08
C14	Cafeteria and Facilities	0.08	-0.18	**0.89**	-0.10	0.09	0.02
C15	Customer Service Training	**0.97**	0.03	-0.02	-0.10	0.00	-0.03
C16	Room and Amenities	-0.04	-0.10	**0.87**	-0.08	0.11	0.01
C17	Diagnosis and Treatment	0.03	**0.86**	-0.10	0.07	0.03	-0.04
C18	Insurance Issues	0.02	0.09	0.12	-0.02	**0.76**	-0.04
C19	Emergency Services	0.22	0.42	0.13	0.22	-0.03	-0.08
C20	Language & Comm. Barriers	**0.60**	0.18	-0.12	-0.14	0.05	0.33
C21	Patient Rest and Comfort	0.22	**0.53**	0.30	0.04	-0.32	-0.02
C22	Staff Responsiveness	**0.98**	-0.12	0.02	0.05	0.00	-0.04
C23	Patient Privacy	-0.02	0.19	-0.07	-0.02	0.21	**0.50**
C24	Accessibility for Indiv. Disabilities	0.02	-0.15	0.14	0.22	-0.10	**0.71**
C25	Medication Issues	-0.14	**0.65**	0.02	-0.02	0.23	0.11

C26	Departmental Coordination	**0.71**	0.06	0.01	0.17	0.07	-0.08
C27	Front Desk Service	**0.85**	-0.10	0.04	0.11	-0.01	-0.01
C28	Punctuality	0.03	-0.07	0.04	**0.76**	-0.12	0.02
C29	Breastfeeding Support	-0.07	0.06	0.06	-0.04	0.00	**0.45**
C30	Hospital Facilities & Food Quality	-0.02	0.18	**0.71**	-0.13	0.04	0.05
C31	Organizational Efficiency	**0.80**	-0.07	0.12	0.12	0.09	-0.16
C32	Doctor Changes and Qualifications	0.06	**0.72**	-0.17	0.01	0.11	0.12
C33	Staff Attitude	**0.98**	0.11	-0.01	-0.31	0.00	0.07
C34	Staff Overload	**0.55**	0.12	-0.06	0.31	-0.03	0.09
C35	Lift and Equipment Issues	-0.05	0.02	**0.46**	0.13	0.05	0.27
C36	Insurance Billing Mistakes	0.01	-0.06	0.05	0.02	**0.83**	0.08
C37	Room Readiness	0.09	0.10	0.38	**0.45**	-0.13	0.02
C38	Workplace Rushing	**0.60**	-0.01	0.00	0.37	-0.05	0.02
C39	Comparative Evaluation	0.12	**0.47**	0.13	-0.10	0.06	0.00
C40	Inefficient & Disorg. Processes	**0.69**	-0.10	0.00	0.32	0.05	0.01
C41	Amenities Adequacy	-0.03	0.03	**0.70**	0.17	0.05	-0.06

Table 3. Factors, Chosen Items, Factor Loadings and Cronbach alpha.

Factor	Items	Factor Loading	Cronbach alpha
Factor 1	Communication Issues	0.95	0.97
	Service Quality and Professionalism	1.00	
	Doctor's Behavior and Tardiness	0.51	
	Customer Service Training	0.97	
	Language and Communication Barriers	0.60	
	Staff Responsiveness	0.98	
	Departmental Coordination	0.71	
	Front Desk Service	0.85	

	Organizational Efficiency	0.80	
	Staff Attitude	0.98	
	Staff Overload	0.55	
	Workplace Rushing	0.60	
	Inefficient and Disorganized Processes	0.69	
Factor 2	Nursing Care	0.77	0.92
	Patient Safety and Hygiene	0.64	
	Electronic Health Info. Management	0.47	
	Diagnosis and Treatment	0.86	
	Patient Rest and Comfort	0.53	
	Medication Issues	0.65	
	Doctor Changes and Qualifications	0.72	
	Comparative Evaluation	0.47	
Factor 3	Facility Maintenance	0.61	0.90
	Cafeteria and Facilities	0.89	
	Room and Amenities	0.87	
	Hospital Facilities and Food Quality	0.71	
	Lift and Equipment Issues	0.46	
	Amenities Adequacy	0.70	
Factor 4	Appointment System Inconsistency	0.42	0.89
	Waiting Time	0.82	
	Test and Result Processing	0.50	
	Punctuality	0.76	
	Room Readiness	0.45	
Factor 5	Financial Concerns	0.89	0.88
	Insurance Issues	0.76	
	Insurance Billing Mistakes	0.83	

Factor 6	Patient Privacy	0.50	0.61
	Accessibility for Disabilities	0.71	
	Breastfeeding Support	0.45	

Table 4. Retained factors and proposed labels.

Factors	Proposed Factor Labels
Factor 1	Service & Communication Effectiveness
Factor 2	Clinical Care & Patient Experience
Factor 3	Facilities & Amenities Quality
Factor 4	Appointment & Patient Flow
Factor 5	Financial & Insurance Management
Factor 6	Patient Rights & Accessibility

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.

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 alpha and interpretability. Adding more factors increases the cumulative explained variance but compromising cronbach alpha on the additional factors and its interpretability. Removing factors on the other hand maintain cronbach 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 (Knetsch 1989)[40], 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 Vamathevan et al. (2019)[41] 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 Williams et al. (2010)[42].

The themes we identified using online review, gpt-4o-mini model, and factor analysis are supported by previous studies. Systematic review by Ferreira et al. (2023)[43] ranked criteria deemed as the most important to evaluate satisfaction in literature (global analysis). The factors inducted from our study are all listed in his systematic review. According to Ferreira et al. (2023), medical care is deemed as most important in literature (34%), communication (31%), doctor's characteristics (28%), accommodations (23%), admission and discharge (13%), nurse characteristics (11%), appointment (8%), environment (8%), medical expenditure (8%), and organization (8%). Ferreira et al. however derive these numbers through literature reviews thus there are overlapping domains within his full list of 56 criteria, almost similar to what we experienced before conducting factor analysis.

Conclusion

This study underscores the high level of patient satisfaction with private hospitals in Selangor, as evidenced by the predominance of positive Google reviews. The robust dataset of 14,938 reviews, coupled with a thorough analysis using the gpt-4o-mini model, reveals key factors influencing patient experiences, including medical care, communication, and organizational efficiency. Our findings align with existing literature, particularly the systematic review by Ferreira et al. (2023), which highlights similar criteria essential for evaluating patient satisfaction.

Overall, the identified factors provide valuable insights for healthcare providers, offering clear areas for improvement to enhance patient experiences. Future research should continue to explore these dimensions, further validating the relationships between the identified factors and patient satisfaction, and potentially expanding the analysis to other regions or healthcare contexts.

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.

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 waiting time is too long[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 serious illness, something emergency[2,3], they are really lack of experienced staff[1,4].	1. Work Load 2. Emergency Services 3. Patient Safety and Hygiene 4. Doctor's Qualification and Doctor's Change
3	On 2nd Feb 2023 this S**a nurse said will check for me for an available appointment for this specialist doctor, whom I want to see. She didn't revert back to me at all[1,2,3,4]. I called again on 28th February this nurse A***a said my appointment was slot on 23rd March @ 5 pm so on 22nd 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 didn't inform me that the date has been postponed to 24th March[1,2] instead she told me is 3 pm so I thought is in 23rd March at 3 pm. Such a private hospital so incompetent and inefficient[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 really need to improve the admission process[2], because I never seen the improvement of this important section after multiple times visited here: 1) To do a 3 seconds Covid test, patient need to wait for 30 minutes although only 3 patients in queue[1] 2) To sign 3 pages admission document, need to wait for 25 minutes. I saw only 2 staffs multitasking and very busy[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, you would find out what extremely lack of and what need urgently improve[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 feel uncomfortable and painful 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 disappointed with Ms. Wong's service from the Refund Department[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. I followed up on my refund status six times in three weeks and didn't receive any response until today[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 nobody answer[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 hospital not really want entertain cash patients, hospital welcome those with medical card[1,3] which they can order all sort of test and able charge without hassle[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 save paper by issuing appointment letters via email or texts[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 CO_2 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 absolutely useless patient management system[1,2,3]. It's always down which means the poor staff have to resort to manual tracking[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 payment[2] and pharmacy[3] the worst service ever! Gotta wait extremely long[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 doesn't give updates, does not reply or just gives one word replies[1,2]. Likes to admit patients and treats them beyond his area of speciality. Doesn't update the family during the course of the admission[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 request many times for the bag of diapers[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 accountability for the number of diapers used per day? This will settle any potential disputes[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

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., Laverty, 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] Cai, Y., Cai, L., Fu, T., Ye, Y. and Zhou, S., 2018. Analysis of Massive Unstructured Data Model Based on Clustering Algorithm. *Academic Journal of Computing & Information Science*, *1*(1), pp.28-35.

[12] Braun, V. and Clarke, V., 2006. Using thematic analysis in psychology. *Qualitative research in psychology*, *3*(2), pp.77-101.

[13] 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.

[14] Merriam, S.B. and Tisdell, E.J., 2015. *Qualitative research: A guide to design and implementation*. John Wiley & Sons.

[15] Tuan, A. and Grandi, S., 2018. Emerging trends in qualitative research: a focus on Social Media. *Mercati e competitività: 4, 2018*, pp.17-26.

[16] 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.

[17] Stammbach, 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*.

[18] Vaswani, A., 2017. Attention is all you need. *Advances in Neural Information Processing Systems*.

[19] 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).

[20] 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.

[21] 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*.

[22] 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.

[23] 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.

[24] 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.

[25] World Health Organization, 2019. Regional action agenda on harnessing e-health for improved health service delivery in the Western Pacific. Chapter 4, p.22.

[26] Ministry of Health. 2023. Health White Paper for Malaysia. Chapter 3, p.32-55.

[27] 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*.

[28] 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*.

[29] Namita, N.R., Comparison between Traditional topic Modeling and Generative AI Topic Modeling.

[30] 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).

[31] Asaad, W.H., Allami, R. and Ali, Y.H., 2023. Fake Review Detection Using Machine Learning. *Revue d'Intelligence Artificielle*, *37*(5).

[32] 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.

[33] 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.

[34] 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.

[35] 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.

[36] 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.

[37] George, D., & Mallery, P. (2003). SPSS for Windows Step by Step: A Simple Guide and Reference. 11.0 Update (4th ed.). Boston: Allyn & Bacon.

[38] Yusoff, M.S.B., 2019. ABC of content validation and content validity index calculation. *Education*

in medicine journal, *11*(2), pp.49-54.

39 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.

40 Knetsch, J.L., 1989. The endowment effect and evidence of nonreversible indifference curves. *The american Economic review*, *79*(5), pp.1277-1284.

41 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.

42 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.

43 Ferreira, D.C., Vieira, I., Pedro, M.I., 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.