



Final Project Report
CareBrew

Members: James Du, Dasha Moskvitina, Dylan Sun, Joe Meng

Teaching Assistant: Jiading Zhu

Course: ECE1786

Word Count: 1979

Penalty: 0%

Code Repository: <https://github.com/ece1786-2024/CareBrew>

Permissions

Below each team member has indicated their permission decision for the recorded presentation, final report, and source code.

James Du

- Permission to post video: **Yes**
- Permission to post final report: **Yes**
- Permission to post source code: **Yes**

Dasha Moskvitina

- Permission to post video: **Yes**
- Permission to post final report: **Yes**
- Permission to post source code: **Yes**

Dylan Sun

- Permission to post video: **Yes**
- Permission to post final report: **Yes**
- Permission to post source code: **Yes**

Joe Meng

- Permission to post video: **Yes**
- Permission to post final report: **Yes**
- Permission to post source code: **Yes**

Based on the responses above, the teaching team has full permission to post the recorded presentation, report, and source code.

Introduction

CareBrew is an interactive training module, deployed on a website, that promotes empathetic conflict resolution skills for customer service professionals in coffee shops. Empathy in customer service is dependent on multiple complex factors, including the employee's response, the customer's perspective, and the scenario itself. Therefore, assessing and teaching empathy requires an adaptive, dynamic model that provides personalized responses and continuous feedback. CareBrew leverages natural language processing to analyze user responses, provide context-sensitive feedback, and foster emotional intelligence through realistic customer service scenarios. This document outlines the background, implementation, and assessment of CareBrew.

Background and Related Work

A review of customer service-related literature shows the importance of quality customer service in how customers perceive a business [1] [2]. Customers value the ability of a service to meet their needs efficiently, effectively, and empathetically [3] [4]. Many articles analyze the effect of empathy in customer service and conclude its importance, detailing the multiple complex factors that affect customer perception of empathy [3] [5]. However, these articles do not propose actionable methods of promoting empathy in customer service. Existing training modules for empathy in customer service are often healthcare industry-specific [6] and fail to effectively train employees to apply these skills in the workplace [7]. This lack of real-world actionable guidance and industry-specific focus limits the effectiveness of existing work in the field of empathetic coffee shop customer service training.

Illustration

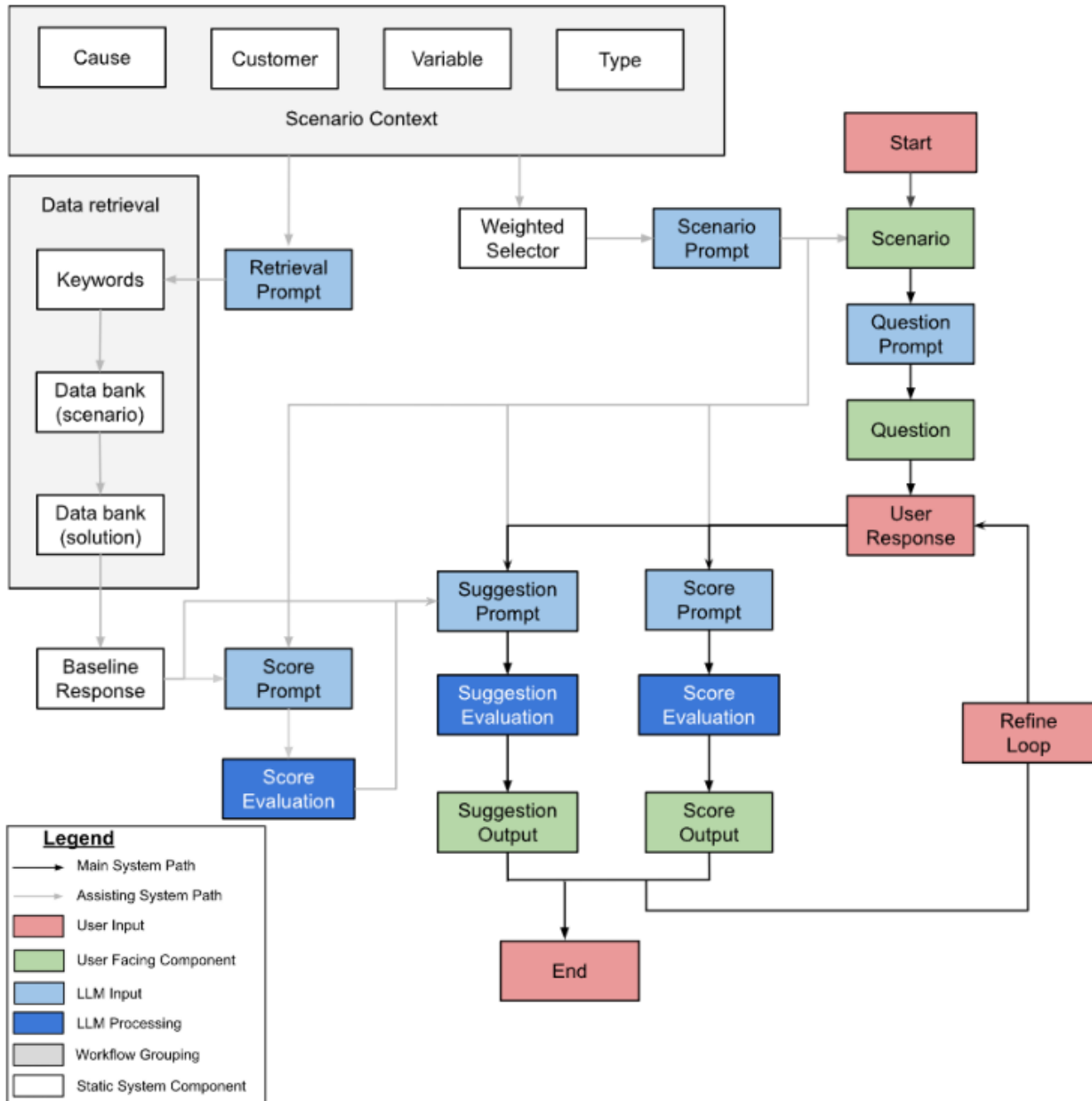


Figure 1: Complete System Diagram

Architecture

Overview

As shown in Figure 1, CareBrew utilizes several custom-prompted GPT-4o agents to train users in empathetic customer service by presenting them with generated coffee shop scenarios and evaluating their responses. This evaluation consists of a score, an explanation, and suggestions for improvement.

Explanation

When the user accesses the training module, a scenario is generated by the *scenario agent*. The prompt utilizes two unweighted factors (cause and customer) and two weighted factors (type and variable) to represent realistic coffee shop customer service scenarios accurately. Based on the scenario, a question is then generated by a *question agent* asking the user how they would respond. Both the scenario and the question are then presented to the user on the CareBrew user interface. The user is then able to type and submit their response on the interface.

The scenario and user response are used as input for the *scoring agent*, which scores the user's response according to a rubric with 5 dimensions of empathetic customer service, given in Appendix A. The score breakdown is then displayed on the CareBrew interface with a justification for each category and an overall score and comment.

Simultaneously, the scenario and user response are used as input for the suggestion generation pipeline. The scenario's keywords are first extracted by the *retrieval agent* and searched for in a database of scenarios, each of which maps to a baseline response. This baseline response is then scored using a modified, shortened *scoring agent*. The scored baseline response, scenario, and user response are used as input for the *suggestion agent*, which generates personalized and scored suggestions for the user based on the categories defined in the *scoring agent* rubric. These suggestions and their corresponding scores are then displayed on the CareBrew interface.

The user is then given the choice to either stop or iterate on their response, which loops the system to start from the user's response again. Additionally, the user can generate a new scenario at any point, which triggers the system to start from the *scenario agent*.

Data and Data Processing

Scenario Generation Process

The scenario generation process was guided by four key factors derived from two sources. A comprehensive survey conducted by Science Korea identified customers' values in coffee shop experiences, providing weighted factors for data generation given in Appendix B [1]. Additionally, one-star Yelp reviews of Starbucks were analyzed to identify customer types and causes of common issues, with the data iteratively summarized into lists. This was combined with a service quality improvement article to formalize the final factors used in the generation prompt [8].

These four factors consisted of two weighted and two randomized elements. Weighted scenario types, like *Trust and Reactivity* or *Taste*, were tied to weighted scenario variables like *fulfilling promises* or *beverage taste*. *Customer Types* and *Issue Causes* are randomized as inputs to the parameterized framework for generating prompts using the GPT-4o API, enabling the creation of detailed baseline scenarios.

Baseline Responses

After generating baseline scenarios, corresponding baseline responses were created. These responses were shorter than the recommended detailed responses in the actual model and served to guide and parameterize more complex prompts. They provided a foundational understanding of good customer service practices when the model retrieves default responses.

For consistency, the refined lists of weighted scenario types were further summarized into five main scenario types with three issue types each given in Appendix C. Each issue type included the issue, a short description, the factors involved (a refactored version of the weighted factors used in scenario generation), and possible causes of the issue. This structured approach allowed GPT-4o to generate 426 responses, representing all possible combinations of customer types, issue types, scenario types, and scenario issues.

With retrieval in mind, the model was designed to categorize scenarios based on detailed issue factors and subtypes, ensuring a nuanced understanding of the context. It was also instructed to identify the primary cause of the issue and determine the associated customer type. This structured approach provided comprehensive coverage and facilitated accurate retrieval of customer service scenarios. Integrating weighted factors and randomized variables, the model effectively captured the complexity and variability inherent in real-world customer interactions.

Comparison

As CareBrew consists of multiple prompted GPT-4o agents, the team measured CareBrew's success through human and automated comparison. Each agent was evaluated based on predefined standards and modified if it did not meet the success metrics established by the team.

For baseline scenario generation, each team member reviewed 30 generated scenarios. Using prior knowledge of standard customer service practices, the team verified that the scenarios were acceptable [8]. Particular attention was given to avoiding excessive repetition, identifying scenarios that appeared more than four times during individual reviews.

For the scoring rubric, a diverse subset of 30 scenarios and responses were scored by each team member. This subset was then input into the scoring agent and the results were compared against the team's average score for each response. The standard deviation between the model and the team's scores was deemed acceptable. A set of 30 generated baseline responses from the data retrieval dataset were then input into the scoring agent and achieved reasonable scores. Using the baseline responses, a set of suggestions was generated and input into the scoring agent, achieving an acceptable score.

Quantitative Results

For baseline response validation, both human and automated evaluations were conducted. The model was tested on 50 scenarios and baseline responses; a subset of 10 were independently graded by each team member based on the rubric given in Appendix A. This was compared against the *scoring agent*, which demonstrated a $\pm 9.1\%$ difference ($\pm 2/22$), indicating that the model performed effectively. A subset of 200 baseline responses from the full set of 426 was then evaluated using GPT-4o prompted with a simplified version of the scoring rubric, resulting in an average score of 19.4 out of 22.

To measure the performance of CareBrew, the team compared human-reviewed responses and model-generated outputs. To evaluate the user responses, the team manually rated a subset using the

same rubric used in the *scoring agent* given in Appendix A and compared it against the model's output. The results demonstrated a $\pm 13.6\%$ difference ($\pm 3/22$), indicating a strong alignment between the model output and human judgment.

To evaluate suggestion quality, a random sample of generated actions was tested using the *scoring agent*. The analysis revealed a $\pm 9.1\%$ difference ($\pm 2/22$) between the model's predictions and the evaluation scores. This result demonstrated the model's reliability and consistency in producing actionable suggestions that aligned closely with human evaluation criteria. These measures illustrate the overall accuracy and effectiveness of CareBrew in meeting its intended objectives.

Qualitative Results

Good Performance

The figures below demonstrate a sample CareBrew user experience.

Scenario: A new customer enters the coffee shop and orders a cappuccino with almond milk, explicitly expressing their trust in the quality of the shop's ingredients and procedures. However, due to a communication-induced error, they are served a drink made with regular milk, conflicting with their trust regarding raw materials. The customer, noticing the mistake upon tasting, becomes visibly upset as this issue not only affects their dietary preference but also challenges their trust in the coffee shop's attention to detail and adherence to procedures.

Figure 2: Sample Scenario

In Figure 2, CareBrew's *scenario agent* generated a realistic coffee shop scenario: a customer becomes upset when they realize they are served a drink with incorrect ingredients. This allowed the user to visualize the real-world example when inputting their response. In this instance, the user said: "I would go to the break room to avoid confrontation".

Clarity	Reliability	Empathy	Relevance	Resolution	Overall Score
1	1	1	1	1	5
The response is clear in indicating the user's plan to avoid the situation; however, it doesn't provide a solution for the customer, as their issue still needs to be addressed.	The user appears unable to handle the situation effectively or correctly, displaying inadequate confidence and not providing any solution.	The user's response shows no empathy towards the customer's frustration or their dietary needs. They do not acknowledge the customer's emotions or provide any comfort or effort to make amends.	The response does not address the customer's concern at all. The customer has a problem with their order, and the user's plan to go to the break room does not provide a solution or indicate an attempt to address the issue.	The issue remains unresolved. The customer received an incorrect order and the user did not provide any steps to either rectify the mistake or offer another solution.	Given the lack of attention to the customer's problem, inadequate response, show of empathy, or steps for resolution, the overall score is low (5/22), equating to a performance evaluation of 22.73%.

Figure 3: Sample Score

In Figure 3, CareBrew's *scoring agent* presented a detailed analysis according to each dimension of the rubric given in Appendix A. The *scoring agent* accurately identified that the response was inadequate because it made no attempt to address the customer's problem and provided justification.

Suggestions for a better response:

- **Category:** Reliability
 - **Suggested improvement:** Assure the customer of your ability to correct the error immediately and you will personally oversee the remake of the drink.
 - **Potential rating:** 4/4
- **Category:** Clarity
 - **Suggested improvement:** Clearly apologize for the mistake and explain specifically how you will fix it, such as remaking the drink with almond milk and ensuring correct procedures are followed.
 - **Potential rating:** 3/3
- **Category:** Empathy
 - **Suggested improvement:** Express understanding of their situation and sincerely apologize for the inconvenience, offering a complimentary item or future discount as a gesture of goodwill.
 - **Potential rating:** 5/5
- **Category:** Relevance
 - **Suggested improvement:** Directly address the issue by acknowledging the specific mistake, confirming the almond milk preference, and providing details on rectifying the order.
 - **Potential rating:** 5/5
- **Category:** Resolution
 - **Suggested improvement:** Offer to remake the beverage with almond milk immediately, provide another drink of choice complimentary, and reassure the customer of preventative measures implemented.
 - **Potential rating:** 5/5

Figure 4: Sample Suggestions

As seen in Figure 4, CareBrew's *suggestion agent* provided suggestions relevant to the scenario and the user's response according to the rubric dimensions. The suggestions focused on acknowledging the customer's frustration and remaking the drink with the correct ingredients. The user then input their follow-up response: "I would say 'I'm sorry for the mistake in your order, I understand your frustration. I'll remake your cappuccino with almond milk right away'".

Clarity	Reliability	Empathy	Relevance	Resolution	Overall Score
3	4	4	5	4	20
The response is specific and easy to understand. It directly addresses the error and clearly outlines the next action.	The user shows complete confidence in handling the situation independently, with clear and accurate action.	The user shows affective empathy by validating the customer's emotions and reframing the issue with a comforting response.	The response fully and directly addresses the customer's concern. The user understands the problem with the order and provides a tailored response to fix it.	The response directly resolves the issue with clear and actionable steps. However, it could have included a method to ensure this doesn't happen again in the future.	Overall, the user handled the situation very well, demonstrating clear communication, strong reliability, and high empathy. The response was relevant, offered a concrete resolution, and resulted in a score of 20 out of 22, or about 91%.

Figure 5: Sample Follow-Up Score

In Figure 5, CareBrew's *scoring agent* presented another detailed analysis. After incorporating the suggestion feedback, the score improved significantly. The suggestions successfully improved the score and the scoring agent accurately identified that the new response demonstrated empathetic customer service in this scenario.

Poor Performance

CareBrew's scoring rubric may inflate scores when user responses lack sufficient detail.

Scenario: A loyal, returning customer named Alex, who visits the coffee shop daily and has a regular order, encounters an issue when a new IT system update incorrectly processes their mobile app order, charging them twice and mixing up their usual espresso for a cappuccino. The type of service challenge is an IT induced error, which undermines Alex's trust in the shop's ability to handle their preferences correctly and respond promptly. A staff member must address Alex's dissatisfaction, keeping in mind the importance of faithfully treating the service customer demanded, and demonstrate reactivity to prevent further erosion of trust. User response: refund					
Latest Grading:					
Clarity	Reliability	Empathy	Relevance	Resolution	Overall Score
2	2	1	4	2	11
Though the user response is simple and short, it doesn't give clear and elaborate information. It just says 'refund', which isn't sufficient to notify the person about what's going on or what steps will be taken next.	The user response implies an intention to resolve the issue, suggesting a possible refund, but it doesn't provide full assurance, the procedure to carry out the refund is not explained properly.	The user response shows no signs of empathy - no acknowledgment of any inconvenience caused, nor any empathetic language used.	The response is relevant to the issue, since a refund can rectify a double charge, but it lacks details and specificity.	The response proposes a resolution (a refund) but no concrete steps or plan for carrying out this resolution are given.	The aggregated score of all dimensions is 11 out of a possible 22, indicating a performance of 50%, reflecting insufficient handling of the customer's situation.

Figure 6: Sample Suboptimal Score

In Figure 6, the user replied to a scenario in which a customer’s order was overcharged with a single word: ‘refund’. While CareBrew accurately identified that it was an insufficient response and scored the user accurately in some categories, it inflated the user’s score in ‘relevance’ and ‘clarity’ by inferring additional details in the user’s response.

Discussion and Learning

There is sufficient evidence to conclude that CareBrew effectively improves the quality of customer service in coffee shops. The model-evaluated scores show clear improvements when users apply system-generated suggestions. Even if the initial user response achieves a high score, CareBrew provides an alternative response suggestion that focuses on different dimensions of the rubric. During quality assurance testing, the resulting data demonstrated a clear linear relationship between the degree of empathy in responses versus the evaluation score.

The original goal of CareBrew remained consistent, however, the system grew increasingly complex as new insights emerged from research and discussion. This growth reflects both the adaptability and the rising standards of the project. However, it also introduced challenges, particularly in maintaining focus on core functions under time constraints. A key takeaway from this process is the need to allocate more time for optimizations to reduce runtime and improve overall efficiency. In response to these challenges, a portion of the codebase should be restructured to streamline its functionality, which will mitigate some of the latency issues.

Individual Contributions

Individual	Contributions
James Du	Lead project workflow with TA Supervisor
	Proposed main project structure
	Performed empathy rubric literature review
	Trained/refined scoring prompt
	Modified refinement loop for prompt compatibility
	UI/UX design
	Added instructions to run the project
Dasha Moskvitina	Initial Implementation of main system architecture
	Implemented question generation
	Trained/refined question prompt
	Implemented suggestion generation

	Trained/refined suggestion prompt
	Adapted suggestion scoring from scoring prompt
	Modified data retrieval for suggestion compatibility
	Modified model response formatting for chat history compatibility
Dylan Sun	Performed scenario literature review
	Implemented scenario generation
	Trained/refined scenario prompt
	Gathered/labeled data retrieval dataset
	Implemented data retrieval (including baseline response)
	Trained/refined keyword extraction
	UI/UX design
Joe Meng	Implemented project homepage
	Implemented refinement loop
	Implemented model response formatting
	Initial implementation of suggestion generation
Everyone	Tested scenario generation
	Tested data retrieval
	Tested rubric scoring
	Tested suggestion generation
	Code reviews/cleanup

References

- [1] C. S. Shin, G. S. Hwang, H. W. Lee, S. R. Cho (2015). The impact of Korean franchise coffee shop service quality and atmosphere on customer satisfaction and loyalty. *East Asian Journal of Business Management*, 5(4), 47–57.
- [2] Q. X. Tran, M. V. Dang, N. Tournois (2020). The Role of Servicescape and Social Interaction toward Customer Service Experience in Coffee Stores: The Case of Vietnam. *International Journal of Culture, Tourism and Hospitality Research*, 14(4), 619–637.
- [3] C.M. Clark, U. M. Murfett, P. S. Rogers, S. Ang (2008). Is Empathy Effective for Customer Service? Evidence from Call Center Interactions. *Journal of Business and Technical Communication*, 27(2), 123-153.
- [4] A. K. Khan, B. A. Metri (2011). Understanding Customers' Service Experience: Review and Research Propositions. *International Journal of Business Environment*, 4(1), 45–62.
- [5] B. R. Lewis, V. W. Mitchell (1990). Defining and Measuring the Quality of Customer Service. *Marketing Intelligence & Planning*, 8(6), 11-17. <https://doi.org/10.1108/02634509310027865>
- [6] M. Lajante, M. D. Prete, B. Sasseville, G. Rouleau, M. P. Gagnon, N. Pelletier (2023). Empathy Training for Service Employees: A Mixed-methods Systematic Review. *PLoS ONE*, 18(8): e0289793.
- [7] J. Shen, C. Tang (2018). How Does Training Improve Customer Service Quality? The Roles of Transfer of Training and Job Satisfaction. *European Management Journal*, 36(6), 708-716.
- [8] J. C. Yuan, H. F. Chang, G. H. Tzeng (2012). Evaluation of Service Quality Continuous Improvement in Coffee Shops. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 25(1), 1-11.
- [9] G. Packard, J. Berger (2021). How Concrete Language Shapes Customer Satisfaction. *Journal of Consumer Research*, 47(5), 787-806.
- [10] M. Yeomans, F. K. Boland, H. K. Collins, N. Abi-Esber, A. W. Brooks (2023). A Practical Guide to Conversation Research: How to Study What People Say to Each Other. *Advances in Methods and Practices in Psychological Science*. 6(4).
- [11] N. Rane, A. Achari, S. P. Choudhary (2023). Enhancing Customer Loyalty through Quality of Service: Effective Strategies to Improve Customer Satisfaction, Experience, Relationship, and Engagement. *International Research Journal of Modernization in Engineering Technology and Science*. 5(5).
- [12] S. Lee, J. Lee, J. Kim, S. Park, H. Kim (2021). A Study on the Impact of Coffee Service Quality on Customer Loyalty. *Journal of Advanced Researches and Reports*. 1(1) 25-32.
- [13] D. Marinova, S. K. Singh, J. Singh (2017). Frontline Problem-solving Effectiveness: A Dynamic Analysis of Verbal and Nonverbal Cues. *Journal of Marketing Research*.

Appendices

The following appendices are provided as supplemental material to enhance the understanding of key aspects mentioned in the report. Readers are encouraged to refer to these sections for additional context and detailed explanations.

Appendix A: Rubric Decomposition

This appendix provides a detailed overview of the empathy evaluation rubric for assessing the quality of customer service interactions across five key dimensions: Clarity, Reliability, Empathy, Relevance, and Resolution. Each dimension is graded on a level supported by a description and illustrative example to ensure clarity and consistency in application.

Dimension	Level	Description	Example
Clarity [9] [10] <i>Measures the staff's ability to communicate clearly and effectively.</i>	1: Unclear Communication	The response is vague, confusing, or overly complex. May use jargon or language difficult for a layperson to understand.	"We'll recalibrate the system to improve operational alignment."
	2: Somewhat Clear Communication	The response provides basic clarity but lacks specificity or uses abstract/general language.	"I'll fix that for you so that it's better."
	3: Clear and Concrete Communication	The response is specific, easy to understand, and avoids ambiguity. It clearly explains the action being taken.	"I'll remake your coffee at the correct temperature."
Reliability [1] [5] [11] <i>Measures the staff's ability to handle the situation confidently and accurately.</i>	1: Inadequate Confidence	The staff appear unsure, provide incomplete or incorrect information, or fail to handle the situation.	"I'm not sure what to do in this case."
	2: Hesitant Confidence	The staff show some uncertainty or defer action to someone else.	"I think I can help, but let me double-check first."
	3: Competent but Verified Handling	The staff handle the issue but seek confirmation from a higher authority.	"Let me check with my manager to confirm this for you."
	4: Complete Confidence	The staff handle the situation fully and independently with clear, decisive, and accurate action.	"I can take care of that for you right now—here's what I'll do."
Empathy [3] [4] [5] [12]	1: No Empathy	The response is dismissive, cold, or ignores the customer's emotions.	Customer: "I'm upset my order is late!" Staff: "I don't know what to tell

<i>Measures the staff's ability to connect emotionally with the customer and address their needs</i>			you."
	2: Minimal Empathy	Acknowledges emotions insincerely or superficially.	"Sorry for the delay, we're busy."
	3: Attentive Empathy	Actively listens, acknowledges emotions, and paraphrases the customer's concerns.	"I understand you've been waiting. I'll make sure your order is prioritized."
	4: Affective Empathy	Validates emotions and reframes the issue with comforting or insightful responses.	"I can imagine how frustrating that must be. I appreciate your patience and will fix this right away."
	5: Cognitive Empathy	Fully aligns with the customer's emotional and practical needs, offering a tailored solution.	"I completely understand your frustration. I'll remake your coffee and offer a discount for the inconvenience."
Relevance [1] [3] [13]	1: Irrelevant Response	The response does not address the customer's concern.	Customer asks about a refund; the response discusses promotions
<i>Measures the staff's ability to address the specific issue at hand.</i>	2: Partially Relevant	The response only partially addresses the issue or is tangential.	Customer asks for a refund; the response mentions store policies but doesn't explain the process.
	3: Relevant but Generic	The response addresses the issue but lacks depth or personalization.	Customer asks for a recommendation, and the staff give a generic answer without asking follow-up questions.
	4: Relevant and Specific	The response directly and thoroughly addresses the issue with appropriate detail.	Customer asks about a product, and the staff provide clear, specific details about it.
	5: Highly Relevant and Tailored	The response fully addresses the unique concern with precision and a tailored approach.	Customer asks for a refund, and the staff provide specific steps for resolution, timing, and follow-up.

Resolution [1] [3] [7] [13] <i>Measures the staff's ability to offer and implement an effective solution.</i>	1: No Resolution	The issue remains unresolved or unaddressed.	Customer asks for a refund, and no solution is provided.
	2: Partial Resolution	The response addresses part of the issue but lacks completeness or clarity.	Customer asks for a refund, and the staff mention the process but provide no next steps.
	3: Basic Resolution	The response resolves the issue but is generic or unclear in some aspects.	Customer asks about an issue, and the response provides a basic, non-tailored solution
	4: Effective Resolution	The response directly resolves the issue with clear and actionable steps	Customer asks about a delay, and the staff provide an estimated delivery time and steps to track the order.
	5: Complete and Tailored Resolution	The response fully resolves the issue with a detailed, personalized solution and ensures customer satisfaction.	Customer asks for a refund, and the staff provide a step-by-step guide, timing, and reassurance of resolution.

Appendix B: Weighted Factors and Variables

This appendix provides a detailed breakdown of the weighted factors and associated variables identified through a comprehensive survey conducted by Science Korea, which explored what customers value most in coffee shop experiences [1]. These factors influence the scenario generation process, ensuring alignment with real-world customer priorities. Each factor is assigned an overall value and is further broken down into specific contributing variables. Additionally, customer and cause types are categorized to provide nuanced insights into the dynamics influencing service quality and outcomes.

1. Trust and Reactivity

- **Overall Value:** 10.873
- **Variables**
 - Faithfully treating service customer demanded: 0.732
 - Responding immediately to customer's question: 0.726
 - Providing rapid service in order of visiting customer: 0.708
 - Keeping promise made by customers: 0.706
 - Workers responding immediately to customer's demands: 0.648
 - Attitude to help customers willingly: 0.644
 - Treating customer's demand accurately: 0.628
 - Trust towards raw materials and cooking procedures: 0.567

2. Assurance and Sympathy

- **Overall Value:** 1.931
- **Variables:**
 - Deep interest towards individual customers: 0.810
 - Accomplishing faithfully to customer's demand: 0.756
 - Carrying out service with sympathy: 0.702
 - Being friendly and thoughtful: 0.642
 - Understanding customer's needs and emotions: 0.642
 - Being safe from all situations such as fire: 0.626
 - Trust towards hygiene and safety: 0.600
 - Skill to make coffee and beverages well: 0.472

3. Taste

- **Overall Value:** 1.483
- **Variables:**
 - Taste of beverages: 0.831
 - Taste of foods excluding coffee and beverages: 0.807
 - Tasty product overall: 0.803
 - Taste of coffee: 0.766

4. Materiality

- **Overall Value:** 1.267
- **Variables:**
 - Facilities and equipment: 0.824
 - Store atmosphere including interior: 0.730
 - Physical Environment: 0.671
 - Attire and appearance of workers: 0.626

5. Customer Types

- Regular customer
- New customer
- Walk-in customer
- Online order customer
- Returning customer

6. Cause Types

- Customer-induced issue
- Communication-induced error
- Employee-induced error
- IT-induced error
- Accidental issue

Appendix C: Refactored Type-Variable Scenario Catalogue

This appendix provides a comprehensive analysis of key customer service challenges, categorized into 5 recurring themes.

1. Delayed Service & Slow Response

- **Issue:** Delayed Service Response
 - **Description:** Customers are waiting too long for their orders to be taken or served.
 - **Factors Involved:**
 - Trust and Reactivity (Responding immediately to customer's questions)
 - Taste (Taste of beverages)
 - **Possible Cause:** Slow response time or lack of urgency from staff.
- **Issue:** Customer Uncertainty About Service Timing
 - **Description:** Customers are unsure when their order will be ready, leading to frustration.
 - **Factors Involved:**
 - Trust and Reactivity (Providing rapid service in order of visiting customer)
 - Materiality (Facilities and equipment)
 - **Possible Cause:** Inefficient systems for communicating order wait times.
- **Issue:** Slow or Inaccurate Order Fulfillment
 - **Description:** Orders are either delayed or wrong, leading to dissatisfaction.
 - **Factors Involved:**
 - Trust and Reactivity (Responding immediately to customer's questions)
 - Materiality (Attire and appearance of workers)
 - **Possible Cause:** Errors in the order-taking process, kitchen or barista mistakes, or understaffing.

2. Staff Interaction & Customer Engagement

- **Issue:** Unfriendly or Unhelpful Staff
 - **Description:** Customers feel that the staff is not friendly or helpful, making them uncomfortable.
 - **Factors Involved:**
 - Assurance and Sympathy (Being friendly and thoughtful)
 - Trust and Reactivity (Attitude to help customers willingly)
 - **Possible Cause:** Staff may be disengaged, not trained to be empathetic, or overworked.
- **Issue:** Staff Ignoring or Not Listening to Customer Requests
 - **Description:** Customers' specific requests or preferences are ignored by staff.
 - **Factors Involved:**
 - Trust and Reactivity (Treating customer's demand accurately)
 - Assurance and Sympathy (Understanding customer's needs and emotions)
 - **Possible Cause:** Miscommunication, inattentive staff, or lack of training.
- **Issue:** Lack of Personalization in Service
 - **Description:** Customers feel that the service lacks a personal touch and doesn't cater to their individual preferences.
 - **Factors Involved:**
 - Assurance and Sympathy (Deep interest in individual customers)
 - Trust and Reactivity (Treating customer's demand accurately)

- **Possible Cause:** Staff fails to engage with customers or doesn't offer tailored recommendations.

3. Food & Beverage Quality

- **Issue:** Inconsistent Quality of Food or Beverage
 - **Description:** The taste or quality of beverages and food items is inconsistent from visit to visit.
 - **Factors Involved:**
 - Taste (Taste of beverages)
 - Materiality (Physical environment)
 - **Possible Cause:** Issues with ingredient quality, inconsistent preparation, or lack of standardization.
- **Issue:** Unpleasant Taste of Beverages or Food
 - **Description:** Customers find the taste of their beverages or food unpleasant, which diminishes their experience.
 - **Factors Involved:**
 - Taste (Taste of beverages)
 - Materiality (Facilities and equipment)
 - **Possible Cause:** Poor ingredient quality, improper preparation, or an error in the recipe.
- **Issue:** Temperature Issues with Coffee or Food
 - **Description:** Customers receive beverages or food that are either too hot or too cold.
 - **Factors Involved:**
 - Taste (Taste of coffee)
 - Trust and Reactivity (Providing rapid service in order of visiting customer)
 - **Possible Cause:** Temperature control issues, equipment malfunction, or poor timing.

4. Physical Environment & Cleanliness

- **Issue:** Dirty or Unappealing Store Environment
 - **Description:** The physical environment of the store (e.g., cleanliness, décor, or maintenance) is uninviting or dirty.
 - **Factors Involved:**
 - Materiality (Store atmosphere including interior)
 - Assurance and Sympathy (Trust towards hygiene and safety)
 - **Possible Cause:** Lack of regular cleaning or outdated equipment and décor.
- **Issue:** Unprofessional or Unkempt Appearance of Staff
 - **Description:** Employees' attire or appearance doesn't meet hygiene or professionalism standards.
 - **Factors Involved:**
 - Materiality (Attire and appearance of workers)
 - Assurance and Sympathy (Being safe from all situations such as fire)
 - **Possible Cause:** Lack of proper uniforms, poor grooming standards, or failure to maintain hygiene protocols.
- **Issue:** Inadequate Allergy or Dietary Information
 - **Description:** The menu doesn't provide enough information about allergens or dietary restrictions (e.g., gluten-free, vegan).

- **Factors Involved:**
 - Assurance and Sympathy (Understanding customer's needs and emotions)
 - Taste (Taste of beverages)
- **Possible Cause:** Lack of clear labeling or communication regarding food allergies.
- **Issue:** Substandard Hygiene or Safety Practices
 - **Description:** Customers feel uncomfortable due to a lack of visible hygiene or safety measures (e.g., staff not wearing gloves, unclean bathrooms).
 - **Factors Involved:**
 - Assurance and Sympathy (Trust towards hygiene and safety)
 - Materiality (Physical environment)
 - **Possible Cause:** Inadequate cleaning schedules or failure to meet safety regulations.

5. Customer Expectations & Service Communication

- **Issue:** Confusing Loyalty Program or Promotions
 - **Description:** Customers have trouble understanding or using the store's loyalty program or promotions.
 - **Factors Involved:**
 - Trust and Reactivity (Keeping promise made by customers)
 - Materiality (Physical environment)
 - **Possible Cause:** Complex or unclear promotion rules, or staff unfamiliar with the program.
- **Issue:** Misleading or Unclear Pricing
 - **Description:** Customers are confused or upset by pricing, especially if the menu does not match the final bill.
 - **Factors Involved:**
 - Materiality (Physical environment)
 - Trust and Reactivity (Keeping promise made by customers)
 - **Possible Cause:** Menu prices not updated or hidden fees on the final bill.
- **Issue:** Unclear Communication of Menu Choices
 - **Description:** Customers are unsure of the menu options or ingredients in their beverages/food.
 - **Factors Involved:**
 - Assurance and Sympathy (Deep interest in individual customers)
 - Materiality (Store atmosphere including interior)
 - **Possible Cause:** Lack of clear signage or insufficient communication from staff.
- **Issue:** Unclear or Unfriendly Return/Refund Process
 - **Description:** Customers encounter frustration when attempting to return or exchange products (e.g., if something is wrong with their order).
 - **Factors Involved:**
 - Assurance and Sympathy (Carrying out service with sympathy)
 - Trust and Reactivity (Responding immediately to customer's questions)
 - **Possible Cause:** Unclear policies or reluctance from staff to address issues.