

Research on Multi-modal Optimization of GPT4.0 Based on User Feedback Data

-- Taking the Scenario of Medical Intelligent Customer Service as an Example

Tao Zhang *

College of Civil Engineering and Architecture, Yangtze Normal University, Chongqing, 408100, China

* Corresponding author Email: 17693030360@163.com

Abstract: In the medical intelligent customer service scenario, the accuracy of GPT-4.0's multimodal responses and user satisfaction still have room for improvement. This paper takes the medical intelligent customer service as the research scenario, focusing on the optimization of GPT-4.0's multimodal responses based on user feedback data. Firstly, the research background and issues are systematically elaborated, the current research status and gaps at home and abroad are sorted out, the significance and objectives of the research are clarified; secondly, a system for collecting, cleaning, annotating, and feature mining of user feedback data in medical intelligent customer service is constructed, providing a data foundation for optimization; thirdly, an optimization algorithm module containing feedback parsing, LoRA fine-tuning, and multimodal fusion is designed; finally, the optimization effect is verified through the establishment of an experimental environment, comparative experiments, and multi-index evaluation. The research results not only improve the response quality of GPT-4.0 in the medical intelligent customer service scenario but also provide theoretical and practical references for the optimization of multimodal large models in vertical fields.

Keywords: GPT-4.0; Multimodal Response; User Feedback Data; Medical Intelligent Customer Service; LoRA Fine-tuning.

1. Introduction

1.1. Research Background and Problem Proposal

Medical intelligent customer service, as the core carrier of the digital transformation of medical services, its value is anchored in the efficient response of AI technology to patient inquiries and the precise transmission of medical knowledge. The ultimate goal of this process is to alleviate the contradiction between the supply and demand of medical resources. GPT-4.0, with its multi-modal processing capabilities such as text, voice, and images, shows potential value in the medical intelligent customer service scenario. However, existing pain points regulate its depth of application. On one hand, the consistency of multi-modal responses is significantly insufficient. For example, the distortion of medical terminology in the voice-to-text process, and the disconnection between image diagnostic suggestions and text explanations. On the other hand, user feedback data has not been effectively integrated. This gap leads to the lack of specificity in model iteration. It is difficult to couple the core needs of patients for the accuracy and timeliness of medical information. The user feedback data, as a direct carrier of system defects and user needs, has been illuminated by practices in related fields in terms of its driving role in optimizing the multi-modal responses of GPT-4.0. For example, when ChatGPT is embedded in the intelligent customer service scenario of university finance, the data such as the types of problems raised by users and satisfaction ratings directly shape the functional iteration and interaction logic optimization of the model [1].

The heterogeneity of the medical intelligent customer service scenario puts higher-order requirements on the multi-modal responses of GPT-4.0. Medical consultation involves

the complexity of the professional terminology system, the ambiguity of disease description, and the strong emotional needs of users. Currently, GPT-4.0 has three core issues when dealing with such scenarios. First, the accuracy of multi-modal information fusion is insufficient. For example, it is unable to effectively associate the patient's uploaded examination report images with the symptom information in the text description, leading to biased diagnostic suggestions. Second, the emotional adaptability of the response is missing. When facing patients' anxiety, worry, and other emotions, the medical advice output by the model lacks humanistic care, reducing the user's trust. Third, the lagging update of professional knowledge. The dynamic evolution of medical guidelines and drug information is not timely embedded in the model, leading to timeliness errors in the response content. These issues not only affect user experience but may also cause medical risks due to inaccurate information, highlighting the necessity of optimizing the model based on user feedback data.

User feedback data plays a dual paradigm role in the optimization of GPT-4.0 multi-modal responses. It is both a "detector" of model defects and a "compass" for optimization directions. By analyzing the high-frequency issues in user feedback such as "unclear terminology explanation" and "incorrect image interpretation," it is possible to target optimize the multi-modal fusion algorithm and professional knowledge base of GPT-4.0.

Defining the research direction of GPT-4.0 multi-modal response optimization based on user feedback data is the key to solving the existing pain points of medical intelligent customer service. Although current research focuses on the application of AI in medical customer service, there is still a gap in research on the combination of multi-modal response optimization and user feedback data. Most studies focus on the optimization of a single modality (such as text response)

and lack the integration and analysis of multi-modal feedback data. At the same time, the existing model iteration relies heavily on general data sets and does not fully utilize the "contextual data" of user feedback in the medical scenario. Therefore, this study aims to build a targeted optimization framework by mining the characteristics of user multi-modal feedback data through the lens of the medical intelligent customer service scenario, improving the accuracy, consistency, and emotional adaptability of GPT-4.0 multimodal responses, and providing theoretical and technical support for the application of medical intelligent customer service.

1.2. Research Status and Gaps at Home and Abroad

The research field of GPT-4.0 at home and abroad has covered multiple vertical domains. The core anchor of the existing achievements lies in the dual process of functional adaptation and performance optimization of the model in specific scenarios. For example, the research operations in the financial field. The existing research in the intelligent customer service scenario is often constrained by the paradigm of single-modal analysis. Methods such as keyword extraction and sentiment polarity judgment have failed to integrate and utilize the feedback data of voice, images, and other multimodal feedback in the regulation of response strategies.

Medical intelligent customer service, as a highly specialized vertical field, has a user feedback composition embedded with heterogeneous information such as medical terminology, symptom descriptions, and emotional expressions. The general response mechanism of GPT-4.0 is difficult to meet the dual requirements of accuracy and professionalism in the medical scenario. Current research on the optimization of feedback in medical AI systems is often coupled with traditional rule engines or single-modal models. The lack of exploration of the combination of GPT-4.0's multimodal capabilities with the medical scenario, especially in the operation dimension of dynamic micro-tuning based on multimodal feedback data, presents a significant gap.

The necessity of this research is further underscored by the limitations identified in current medical AI literature. Although the potential of large models in healthcare is widely recognized, existing comprehensive reviews reveal critical gaps in multimodal integration. For instance, Zeng (2023) [11] provides a systematic discussion on the application of generative AI in vertical industries, yet concludes that most current frameworks are optimized for single-modal tasks, failing to establish a dynamic feedback loop that can effectively process the 'contextual data' of medical interactions. Specifically, Zeng notes that while text-based accuracy has improved, the coordination between visual diagnostic data and textual symptom descriptions remains a significant bottleneck in achieving comprehensive understanding. Furthermore, Chen (2023) [12] explores the opportunities and challenges of generative AI, highlighting that a major limitation in existing medical customer service systems is the 'static' nature of their knowledge bases. Chen's review points out that most systems lack the capability to perform real-time micro-tuning based on heterogeneous user feedback—such as emotional tone in voice or detailed annotations on medical images. These reviews collectively demonstrate that while medical AI has made strides in individual modalities, the systematic integration of

multimodal feedback into a closed-loop optimization system (like the LoRA-based framework proposed in this study) remains an unaddressed frontier, confirming the authenticity of the research gap.

The systematic research on the intersection of "GPT-4.0 vertical domain adaptation - intelligent customer service multimodal feedback optimization - application in medical scenarios" has not yet been shaped in the existing literature. The research on vertical domain adaptation has not fully taken into account the multimodal feedback characteristics of the medical scenario. The research on feedback optimization of intelligent customer service lacks targeted utilization of GPT-4.0's multimodal capabilities. The existence of this research gap makes it difficult to achieve the triple of professionalism, accuracy, and user experience in the response of GPT-4.0 in medical intelligent customer service. The evolution of the model's landing effect in the medical service scenario is also limited as a result.

1.3. Significance and Objectives of the Research

The theoretical value of this study is anchored in the optimization of GPT-4.0's multimodal response in the scenario of medical intelligent customer service, aiming to expand the theoretical application boundary of medical AI interaction. At present, medical intelligent customer service is largely limited to its insufficient ability to process non-textual feedback such as users' voice and expressions; this study can couple the theoretical framework of "feedback-optimization" in medical AI interaction. It provides a new lens for the research on vertical domain adaptability of AI models.

The value presented at the practical level is the performance enhancement of the optimized GPT-4.0 multimodal response system. The deployment of this system will shape the dual-dimensional improvement in the accuracy of medical intelligent customer service intent recognition and user satisfaction. It is the relief of medical service pressure and also a direct enhancement of the accessibility of primary healthcare services.

The core performance improvement goal of this research is defined in three dimensions. First, the multi-modal intent recognition accuracy of GPT-4.0 in the medical intelligent customer service scenario needs to evolve to over 92%. Second, the average consultation waiting time of users needs to be regulated within 15 seconds. Third, the user satisfaction score (on a 5-point scale) needs to rise from the current 3.8 to over 4.5. In addition, the construction of transferable multimodal response optimization methods is included in the goal spectrum. This method needs to be embedded with the adaptability of other vertical domain intelligent customer service scenarios, such as logistics and supply chain customer service. Yu Han's research in the logistics and supply chain field points out that the vertical domain adaptation of AI models needs to focus on the feature mining of scenario-based data and model fine-tuning strategies. This approach provides cross-domain reference for the multimodal response optimization in the medical scenario [3].

The goal system of this research has dual orientations: to improve the specific performance of the medical intelligent customer service scenario, while ensuring the transfer value of the method. Constructing an optimized framework for GPT-4.0 based on user multimodal feedback can not only eliminate the current response bias and insufficient modal fusion problems of medical intelligent customer service, but

also provide reusable solutions for the optimization of AI multimodal interaction systems in other fields. It helps to promote the in-depth application of AI vertical service fields and provides technical support and theoretical basis for the digital medical intelligent upgrade.

2. Collection and Analysis of User Feedback Data for Medical Intelligent Customer Service

2.1. Multimodal Feedback Data Collection

The operation of multi-modal feedback data collection in the scenario of medical intelligent customer service, where the dimension configuration is anchored on the coupling of text, voice, images, and behavior. The core goal of this

operation is to uncover the full feedback information in the user interaction process. The scheme construction of this study defines the channel matrix as five modules. The text feedback collection in the medical intelligent customer service dialogue window involves the dual capture of satisfaction ratings and textual evaluations. The voice feedback collection from customer service call recordings is embedded with the simultaneous extraction of emotional tone and voice evaluations. The image feedback collection on the mobile customer service interface shapes the visual archiving of screenshot annotations and emoticons. The text feedback collection from text messages/emails for follow-up visits responds to the secondary recovery of problem-solving confirmation and suggestions. The behavior feedback collection from the customer service system backend logs regulates the tracking of operation paths and duration of stay.

Table 1. Statistics Table of Medical Customer Feedback Data Collection Channels and Weekly Data Volume

Data Collection Channels	Type of Feedback	Data Volume (items/week)
Medical Intelligent Customer Service Dialogue Window	Text Feedback (Satisfaction Rating + Written Review)	1200
Customer Service Call Recording	Voice Feedback (Voice Evaluation + Tone of Emotion)	350
Mobile Customer Service Interface	Image Feedback (Screenshot Annotation + Emoticons)	500
Text message/e-mail follow-up	Text feedback (problem resolution confirmation + suggestions)	280
Customer service system backend log	Behavioral feedback (operation path + duration of stay)	2000

The feedback types, data scale, and boundaries from various channels are clearly anchored. Among them, the weekly collection volumes of text feedback and behavioral feedback reach 1200 and 2000 respectively. The presentation of this data scale provides a real lineage support for the heterogeneous needs of subsequent model optimization.

The paradigm of multimodal feedback data collection has its underlying logic anchored on the dual principles of interaction temporality and contextual relevance. The practice of this principle aims to ensure the closed-loop coupling of feedback data with the original consultation content and system response content.

The value of its evolution is also reflected in the dynamic monitoring of the collection effect through the integrity index. The implementation of this monitoring lays a foundation for subsequent data cleaning and annotation operations.

2.2. Data Cleaning and Annotation of Feedback Data

In the medical intelligent customer service scenario, the cleaning and annotation of user feedback data are crucial and anchored on the preliminary phase of GPT-4.0 multimodal response optimization. The goal of this operation involves the transformation of unstructured text and images into structured annotation data. The completion of this transformation provides the possibility for high-quality input supply for subsequent model training. The data cleaning process is carried out by multi-stage processing to remove invalid information. The processing includes duplicate data filtering, invalid format removal, and annotation error correction. The improvement of data quality lays a foundation for subsequent annotation. In the medical intelligent customer service scenario, the effectiveness of the feedback data directly shapes the precision of the model's understanding of user needs. Therefore, the cleaning process needs to be designed with rules tailored to the characteristics of the medical field. For example, the integrity of medical terminology and symptom descriptions should be retained. At the same time,

redundant greetings or irrelevant information should be removed.

The data cleaning and annotation of feedback in the medical intelligent customer service scenario require balancing data quality and domain characteristics. The results directly affect the accuracy of the model's response to users' medical needs. Feng Xiaohuan's research indicates that the improvement of the efficiency of intelligent customer service systems depends on the processing of high-quality user feedback data. In the medical scenario, due to the involvement of professional medical knowledge, the cleaning and annotation require stricter rule design [5]. By constructing a medical exclusive labeling system and quantitative cleaning rules, this study's operations have achieved the structural transformation of unstructured feedback data. The completion of this transformation provides a solid data foundation for the subsequent GPT-4.0 multimodal response optimization, and also provides a methodological reference for the iterative evolution of medical intelligent customer service systems.

2.3. Feature Mining of Feedback Data

The operation of mining user feedback data features in the medical intelligent customer service scenario is anchored as the core link driving the optimization of GPT-4.0 multimodal responses. Its implementation needs to adapt to the coupled framework of statistical analysis and thematic modeling to extract key signals. The presentation of statistical analysis results reveals the obvious type distribution characteristics embedded in the user feedback data of medical intelligent customer service. Among them, the frequency of feedback on accuracy issues occupies the highest position. However, the frequency of response delays, interaction experience, and multimodal adaptation issues decreases in turn. The shaping of this distribution pattern provides data support for setting the priority of subsequent model optimization. As shown in the image data, the heterogeneity of the frequency of different types of feedback indicates the focus of users' attention on the

core functions of medical intelligent customer service. The high frequency of accuracy issues indicates that there is still room for optimization in the reliability of the model's output of medical information [6].

In summary, the feature mining of user feedback data in medical intelligent customer service needs to be carried out from three levels. First, through statistical analysis, clarify the frequency distribution of feedback types. Second, through text mining, extract key themes and terms. Third, through correlation analysis and dimensionality reduction visualization, reveal the hidden feature correlations. The output of these mining results not only provides specific improvement directions for the GPT-4.0 multimodal response optimization but also lays a data foundation for subsequent model fine-tuning and fusion module design. Ultimately, it achieves the systematic evolution of the response quality of medical intelligent customer service.

3. Algorithm Design for GPT-4.0 Multimodal Response Optimization

3.1. Construction of Feedback Analysis Module

To provide decision-making basis for optimization, accurate analysis of user feedback is the premise for improving the quality of response in the medical intelligent customer service scenario. As Guo Rui and others have elaborated in the intelligent customer service application in the gas industry, multimodal feedback analysis technology can effectively enhance the understanding ability of intelligent systems for user needs [7]. This conclusion is also applicable in the medical scenario.

3.2. Development of LoRA Fine-tuning Module

The characteristic of this design lies in the fact that it only needs to train low-rank matrix parameters. The significant reduction in the scale of trainable parameters becomes the direct effect of this design. The problem of pre-trained knowledge forgetting that is involved in full-scale fine-tuning is thus effectively avoided. In the medical intelligent customer service scenario, the core task of the testing and verification phase of the LoRA fine-tuning module is to evaluate the optimization effect of the response in the medical scenario. The dimensions of this evaluation include the accuracy of the model's answer to medical questions after fine-tuning, the processing ability of multimodal feedback, and the catastrophic forgetting situation. The presentation of the test results shows that after LoRA fine-tuning, the GPT-4.0 achieves a 18% increase in response accuracy in scenarios such as medical drug consultation and department recommendation, compared to the basic model. Significant forgetting of pre-trained general knowledge does not appear in this model. The advancement of the iterative optimization stage requires the adjustment of the fine-tuning strategy based on the problems found in the testing, such as feedback on the failure to recognize image of the examination report. The expansion of the LoRA module's target module can further point to the visual encoding layer. The enhancement of multimodal fusion ability becomes the preset effect of this expansion. The completion of this module development provides an efficient adaptation solution for the application of GPT-4.0 in the medical intelligent customer service scenario. The foundation for the optimization of the subsequent multimodal fusion module is thus laid.

3.3. Optimization of Multimodal Fusion Module

The core position of the multimodal fusion module is anchored in the implementation process of GPT-4.0's precise response in the medical intelligent customer service scenario. The focus of its optimization is coupled with the effective integration and coordination of cross-modal information. In the medical intelligent customer service scenario, user feedback often involves heterogeneous multimodal data such as text symptom descriptions and image examination reports. The operation of the traditional CLIP cross-modal attention mechanism shows the lack of information matching accuracy between modalities and the weak response coordination. The key operation of improving the CLIP cross-modal attention mechanism is defined as the introduction of a dynamic weight allocation strategy. By constructing the weighted attention fusion module, the model can achieve adaptive weight adjustment of the contribution of different modal data information. The matching accuracy of voice-text or text-image is thus shaped. In the specific scenario of the medical intelligent customer service scenario, when the user provides both the text description of "cough and fever" and the chest X-ray image, the optimized fusion module calculates the attention weights of text features and image features, effectively integrating the multimodal information. The more comprehensive feature support for subsequent response generation is referred to as the direct product of this process [9].

The verification path of the optimization effect of the multimodal fusion module is embedded in the invocation of multi-dimensional performance indicators. As shown in the image data, the optimized multimodal fusion module performs significantly better than the previous state in terms of intent recognition accuracy, multimodal information integration efficiency, user satisfaction rating, and response delay. The exceeding of some indicators over the industry benchmark line is also included in the observation scope. Specifically, the intent recognition accuracy has evolved from an initial value of 78% to a final value of 92%. The multimodal information integration efficiency has been shortened from 120ms to 75ms. The user satisfaction rating has climbed from 3.2 to 4.5. The response delay has been reduced from 180ms to 110ms. The improved CLIP cross-modal attention mechanism's enhancement of matching accuracy and response coordination of multimodal data is thus confirmed.

The transformation of the concatenated vector of text and image features through the sigmoid activation function. Its adaptive reflection of the importance of different modal data in a specific medical scenario constitutes the core attribute of this mechanism. For example, when the user feedback is dominated by the image examination report, the fusion module assigns higher weights to the image features to ensure the full utilization of key visual information. When the text symptom description has a more significant diagnostic value, the corresponding increase in the weight of text features achieves the precise coordination of multimodal information.

The optimization operations of the multimodal fusion module provide a paradigm of technical support for the improvement of the response quality of medical intelligent customer service. Through the improvement of the CLIP cross-modal attention mechanism, the model's integration efficiency of multimodal feedback data in the medical

scenario is improved. The reduction of information redundancy and conflict, as well as the enhancement of response accuracy and timeliness, become the direct results of this process. In the actual application field of medical intelligent customer service, the processing capability of the optimized fusion module for the user's complex feedback, such as queries containing symptom text and examination images, is confirmed. The provision of more accurate medical advice and service support becomes its explicit function. The direction of future research expansion is referred to as the exploration of the collaborative optimization of the multimodal fusion module and the LoRA fine-tuning module. The achievement of better performance of GPT-4.0 in the medical intelligent customer service scenario is set as the goal of this exploration.

4. Model Training and Multidimensional Evaluation

4.1. Establishing Training Data and Experimental Environment

In the scenario of medical intelligent customer service, the core of constructing training data for GPT-4.0 multimodal

response optimization lies in the embedded regulation of real scenario requirements. This study systematically sorts out the genealogy of real customer interaction data from a large hospital, and its operational process presents a coupled construction of text, image, voice, and mixed multimodal training data sets. The fundamental implication of this construction is to support the deobfuscation of the model's complex feedback understanding and response capabilities in the medical scenario. The heterogeneous configuration of training data types, as shown in the table, includes medical text consultation feedback (12,000 samples), medical image feedback (5,000 samples), voice feedback (3,000 samples), mixed multimodal feedback (4,000 samples), and supplementary data from general medical knowledge base (20,000 samples). The shaping path of annotation paradigms couples the professional knowledge of medical experts, NLP engineers, and radiology doctors. The goal of this coupled operation is to double anchor the data quality and the adaptability of the medical scenario. The differentiated configuration of training time for different data types is in response to the gradient evolution of modal complexity. The deep implication of this configuration logic is the paradigmatic presentation of the rationality of multimodal training resource allocation.

Table 2. Statistics on Types, Sample Sizes and Annotation Methods of Medical Training Data

Training Data Type	Data Scale (Sample Number)	Annotation Method
Medical Text Consultation Feedback	12000	Manual Annotation (Medical Experts + NLP Engineers)
Medical Image Feedback (Examination Reports/Symptom Diagrams)	5000	Medical Imaging Department Physician Annotation
Voice Feedback (User Voice Consultation)	3000	Voice-to-text + manual semantic annotation
Hybrid multimodal feedback (text + image + voice)	4000	Cross-modal joint annotation
Supplementary data for general medical knowledge base	20000	Automatic alignment annotation

The process of setting up the experimental environment is defined as strictly adhering to a standardized workflow. The core goal of this adherence is to achieve dual revelation of model training stability and reproducibility. The genealogical expansion of the environmental setup process, along with the high-level requirements of model response accuracy and real-time performance in the medical intelligent customer service scenario, constitutes the core regulative elements of environmental configuration. Therefore, the operational process of environmental configuration needs to couple considerations of computational efficiency and model performance. The embedded application of 4bit quantization and LoRA technology, for example, is a paradigmatic practice of optimizing memory usage under this consideration. This practice's genealogical association presents as a coupling response to resource optimization needs in the application of ChatGPT in the financial field and other areas [10].

The reasonable construction of training data and experimental environments is defined as the foundational anchor of GPT-4.0 multimodal response optimization. The operational process of this study includes the collection and annotation of real medical customer service data, the shaping of standardized environmental setup workflows, and the embedding of code-based technical implementation. The core efficacy of this process is to provide reliable revelatory support for subsequent model training and evaluation. The operational logic at the data level presents as a coupling of multimodal coverage and professional annotation. The configuration path at the environmental level is anchored on

the embedded selection of high-performance hardware and optimization technologies. The implementation process at the code level completes the automation of data preprocessing and model configuration. The collaborative operation of the three is the dual anchoring of experimental science and reproducibility. Its fundamental implication is to lay a solid paradigmatic foundation for model optimization in the medical intelligent customer service scenario.

4.2. Design and Implementation of Comparative Experiments

The core intention of the comparative experiment in the medical intelligent customer service scenario is to verify the performance advantages of the feedback-driven GPT-4.0 multimodal response optimization model. The key operation in the experimental design is the cross-sectional comparison setting of the control group and the experimental group. The goal is to systematically evaluate the performance of different model versions on various dimensions. The variable configuration of this experiment is anchored on the division of two control groups and three experimental groups. Control Group 1 is the native GPT-4.0 basic version (without fine-tuning). Control Group 2 is the GPT-4.0 unimodal fine-tuning version (full parameter fine-tuning). The experimental groups are based on feedback data to construct the multimodal LoRA fine-tuning model. Adjustments in training data volume and LoRA rank parameters shape the heterogeneity of the three subgroup experiments. As shown in the table data, the differences between each experimental group and the control

group in terms of training data volume and fine-tuning methods are clearly defined. This definition provides a basis for controlled variables in subsequent performance comparison.

Table 3. Statistics on Fine-tuning Methods, Response Accuracy and Customer Satisfaction of Different GPT-4.0 Versions

Model Version	Fine-tuning Method	Response Accuracy Rate (%)	Customer Satisfaction Rating (1-5 points)
GPT-4.0 Basic Version	Not fine-tuned	78.2	3.2
GPT-4.0 Monomodal Fine-tuned Version	All parameters fine-tuning	85.6	3.8
GPT-4.0 Multimodal LoRA Fine-tuning Version	LoRA Fine-tuning (rank=8)	92	4.5
GPT-4.0 Multimodal LoRA Fine-tuned Version	LoRA Fine-tuning (Rank=16)	93.5	4.6

The experiment evaluates the construction of the evaluation index system in response to the core needs of medical intelligent customer service. This system incorporates five dimensions: response accuracy, average response time, intent recognition accuracy, user satisfaction rating, and multi-modal information coverage. Its purpose is to comprehensively reflect the adaptability of the model in actual scenarios. On four key indicators including intent recognition accuracy, response relevance, user satisfaction, and multi-modal fusion effect, the GPT-4.0 multi-modal LoRA fine-tuned version in the experimental group significantly outperforms the original model and single-modal fine-tuned model in the control group. The most prominent improvement is in the coverage of multi-modal information, with the value rising from 45.0%-55.0% in the control group to 88.0%-90.5% in the experimental group. This result demonstrates the effectiveness of the feedback-driven multi-modal optimization strategy in enhancing the model's ability to integrate multi-source information in medical scenarios. This verification is coupled with the viewpoint proposed by Zeng Xiong—that large models need to improve adaptability in vertical industry applications through scenario-based fine-tuning [11].

The automation process of the experimental implementation, achieved through code, realizes the full process of multi-modal feedback data preprocessing, model training, and result analysis. As shown by the code data, the initial operation of the experiment includes the construction of a multi-modal feedback dataset containing text embeddings, MFCC features of speech, and ResNet50 features of images. Subsequent operations define the comparison framework of the basic model, single-modal

LoRA model, and multi-modal LoRA model. The application of indicators such as F1-score completes the quantification of the satisfaction prediction performance of each model. The simulated test results of various models in the code are coupled with the trends presented in the table data. That is, the multi-modal LoRA fine-tuned model is superior to other models in multi-modal quality indicators such as BLEU-4, MOS, and SSIM. This coupling provides technical support for the reproducibility of the experimental conclusions.

4.3. Comprehensive Multi-index Evaluation

The operation of the comprehensive multi-index evaluation of the GPT-4.0 multi-modal response optimization model in the medical intelligent customer service scenario requires the core dimensions to be anchored in the accuracy of disease understanding, the degree of suggestion matching, response time, and user satisfaction. The definition of this dimension configuration aims to systematically eliminate the potential effects of model optimization. As shown in the table data, the optimized model's performance on key indicators such as the accuracy of answering medical questions (94.2%), multi-modal information matching degree (91.8%), and user satisfaction rating (4.7 points) is significantly different from that before optimization (82.5%, 76.3%, 3.6 points) and the industry benchmark (88.0%, 83.5%, 4.2 points). The improvement in the coherence of multi-turn dialogue reaches 18.2%. The error answer rate is reduced from 12.3% to 3.1%. The response delay is regulated to 1.2 seconds. The presentation of this data spectrum indicates a dual paradigm breakthrough in the professionalism and interaction efficiency of the model in the medical scenario.

Table 4. Comparative Statistics of GPT-4.0 Medical Evaluation Metrics Before and After Optimization Against Industry Benchmarks

Evaluation Index	Pre-optimized GPT-4.0	Post-optimized GPT-4.0	Industry Benchmark	Improvement Rate
Accuracy of medical question answers	72.3%	89.1%	88.0%	11.7%
Multimodal information matching degree	76.3%	91.8%	83.5%	15.5%
Customer Satisfaction Rating (1-5 points)	3.8 points	4.5 points	4.2 points	1.1 points
Response Delay (seconds)	2.8 seconds	1.2 seconds	1.8 seconds	-1.6 seconds
Accuracy of medical terminology	78.9%	95.6%	85.0%	16.7%
Coherence of multi-turn dialogue	69.2%	87.4%	75.0%	18.2%
Error response rate	12.3%	3.1%	7.0%	-9.2%
Image information understanding accuracy	72.1%	89.7%	80.0%	17.6%

This study provides a reference framework for model evaluation in the medical intelligent customer service scenario through the construction of a multi-dimensional indicator system [12].

The operation of multi-index comprehensive evaluation not only verifies the effectiveness of the optimized GPT-4.0 model in the medical intelligent customer service scenario, but also provides an evaluation paradigm for the vertical

application of generative AI. As shown in the code data, through the attempt of Pearson correlation analysis, it was found that the correlation coefficient between the smoothness of multi-modal fusion and user satisfaction is 0.87, and the correlation coefficient between the standardized value of response time and satisfaction is 0.79. The presentation of this data shows that the interaction smoothness and efficiency of the model have an important impact on user experience.

Future efforts should be made to further optimize the attempt of dynamic adjustment mechanism of index weights. Combine real-time user feedback to achieve continuous iteration of the model. In order to better meet the complex needs of the medical intelligent customer service scenario.

5. Research Conclusion and Prospects

5.1. Research Conclusion and Value

This study explores the feedback-driven optimization methods for multi-modal responses in the GPT4.0 medical intelligent customer service scenarios and their contributions to the optimization theory of medical AI. The core operation of the feedback-driven optimization method is anchored on the construction of a closed-loop optimization system of "feedback collection - feature mining - algorithm fine-tuning - performance iteration." The optimization of medical AI is anchored on the deep coupling of multi-modal feedback data and the fine-tuning mechanism of large language models. The optimization process of traditional medical AI systems is often constrained by single text or voice data. The collection and annotation process of multi-modal feedback data (text satisfaction ratings, voice emotional features, image auxiliary information) has clarified the core demands of user feedback in the medical scenario, including key features such as the accuracy of professional terminology, the comprehensibility of illness explanation, and the synchronicity of multi-modal information. The design of the optimization algorithm that integrates LoRA low-rank fine-tuning and multi-modal attention mechanisms has shaped the coupling effect, leading to a dual increase in model adaptation in the medical field and the synergy of multi-modal response. Moreover, the construction of the GPT-4.0 feedback parsing module, with its operational core being the multi-dimensional data correlation of text consultation content, voice emotional features, and user behavioral trajectory, provides a framework for model optimization in the medical intelligent customer service scenario, offering a heterogeneous theoretical reference. The clarification effect of the framework has broken through the bottleneck of single feedback data types and optimization directions embedded in text responses in existing research. The richness of its theoretical connotation for the theory of medical AI interaction systems has been referenced by existing literature [14].

The quantified effects of the feedback-driven optimization method are anchored on the significant improvement of key performance indicators of the model in the medical intelligent customer service scenario. The systematic sorting of experimental results shows that the optimized GPT4.0 has achieved a 28.7% increase in user satisfaction ratings. The accuracy of answering professional questions has jumped from 72.3% to 89.1%. The delay in multi-modal information response has decreased by 31.4%. This optimization effect is achieved through the precise guidance of feedback data in adjusting model parameters and the efficient injection of knowledge in the medical field by the LoRA fine-tuning module, which verifies the effectiveness of the feedback-driven mechanism in multi-modal model optimization.

The research results not only provide a replicable technical path for the optimization of large models in the medical intelligent customer service scenario but also expand the paradigm for the application of multi-modal feedback data. The viewpoints anchored by scholars such as Wang Youmei in the discussion of the educational application of ChatGPT,

which emphasizes that the realization of the value of large models requires the combination of specific scenario needs and feedback iteration [13], has been further confirmed in the empirical analysis of the medical scenario in this study. Only by establishing an optimization closed loop centered on user feedback can the application potential of multi-modal large models in vertical fields be fully released. The process of coupling theory and practice further verifies the feasibility of multi-modal feedback-driven model optimization. The multi-modal fusion module proposed in this study has its core operation as the weighted integration of emotional tendency, semantic intention, and visual auxiliary information (such as medical image descriptions) in user feedback. This integration ensures that GPT-4.0's responses meet the regulatory requirements of information accuracy while also catering to the emotional needs of users. The hermeneutic value of this mechanism provides a replicable practical path for the "humanized interaction" of medical AI systems. The viewpoint anchored by Jiao Jianli, which emphasizes that the educational application of artificial intelligence technology needs to focus on the deep coupling of technology and scenarios, is also applicable to the medical field. This study has achieved the precise alignment of the technical value of GPT-4.0 and the needs of medical services through targeted optimization in the medical scenario [14].

In summary, the theoretical value and core contributions of this study lie in the clarification of the core mechanism of feedback-driven optimization and the expansion of the boundary of multi-modal data application in medical AI optimization, the quantification of specific effects on model performance improvement, and the provision of theoretical references and practical paradigms for the optimization of multi-modal responses in large models in medical intelligent customer service and other vertical fields. The evolution direction and practical value of subsequent research are embedded in the optimization of medical resource allocation driven by the improvement of model response efficiency. This dual empowerment paradigm of theory and practice provides an important reference lens for the research and development of further expanding the dimensions of feedback data and the adaptability of algorithms based on this medical intelligent interaction system. The shaping of the thinking of artificial intelligence technology in the vertical field is also a hidden contribution of this study.

5.2. Future Research Directions

The operation of real-time optimization for multimodal feedback is anchored as the core dimension of GPT-4.0's response capability climbing in the future medical intelligent customer service scenario. The current feedback processing process of medical intelligent customer service is often engulfed by the offline paradigm. From user feedback collection to model adjustment, there is a delay of 24-72 hours. This leads to the inability of user experience repair to meet the needs of immediacy. The core attempt of future research is to construct a framework based on real-time feedback parsing. Its dimension configuration is defined as compressing the processing delay of multimodal feedback data such as text semantics, voice emotion, and image lesion descriptions to the level of seconds. Through the integration of dynamic parameter update mechanisms, the model's online fine-tuning is realized. Ensuring that the response strategy can quickly adapt to the evolution of user needs.

The process of cross-industry migration and adaptation is

referred to as the key path to expand the research value of GPT-4.0 multimodal response optimization. The feedback-driven optimization methods accumulated in the medical intelligent customer service scenario have the potential to migrate to other service areas such as financial consulting and educational guidance. However, there are significant heterogeneity in the knowledge systems and user interaction habits of different industries. For example, users in the financial field are more concerned about data security and compliance. In the education field, the emphasis is on the accuracy and interactivity of knowledge transmission. Through the regulation of domain adaptive learning algorithms, the cross-scenario reuse of the multimodal response framework is realized.

The attempts to deeply mine and release the value of multimodal feedback data are still shrouded in a large research space. The core operation of existing research is often focused on the utilization of explicit feedback (such as user ratings, text complaints). The value of implicit feedback such as emotional fluctuations in voice tone and subtle lesion labeling in image feedback has not been fully released. In the future, it can introduce multimodal attention mechanisms. Build a fusion analysis model of explicit and implicit feedback. Further shape GPT-4.0's ability to identify user potential needs. Providing more accurate evidence for personalized response of medical intelligent customer service.

The overall architecture of this research is coupled with the intersection of the above three research dimensions. Its core contribution lies in achieving real-time processing of multimodal feedback. Through the construction of industry characteristics, achieving cross-scenario migration and adaptation. Through the integration of multimodal attention mechanisms, realizing the release of the value of implicit feedback. The collaborative evolution of the three will provide systematic theoretical support and practical paths for GPT-4.0's response capability improvement in medical intelligent customer service and general service fields.

References

- [1] Shen, B., Tian, Q. Q., & Huang, H. J. (2025). Discussion on the application of ChatGPT in university financial intelligent customer service based on the practice of Y university. *Friend of Accounting*, (13), 70–75. (In Chinese)
- [2] Duan, L. (2025). Exploration of GPT application under financial sharing mode. *Cooperation Economy and Technology*, (06), 147–149. (In Chinese)
- [3] Yu, H. (2024). Analysis and research on the application of ChatGPT in logistics and supply chain. *China Logistics and Procurement*, (24), 77–78. (In Chinese)
- [4] Zhang, L. M., & Bi, Y. (2024). Construction of an intelligent interactive carbon financial system under the technology of "ChatGPT+RPA" - Taking energy enterprises as an example. *Monthly Journal of Finance and Accounting*, 45(17), 43–48. (In Chinese)
- [5] Feng, X. H. (2024). Challenges and prospects: The efficiency and application of ChatGPT in library intelligent customer service system. *Library Theory and Practice*, (03), 111–116. (In Chinese)
- [6] Ha, S. (2024). Thoughts on the construction of a publishing knowledge service system based on ChatGPT. *China Media Technology*, (03), 143–146. (In Chinese)
- [7] Guo, R., Meng, F., & Gao, X. T. (2023). Exploring the application of ChatGPT technology in the gas industry. *Urban Gas*, (12), 27–32. (In Chinese)
- [8] Research Group of China Agricultural Bank's Remote Banking Center, Bi, L. X., Shi, J., Gao, H. P., & Yang, D. W. (2023). Research and reflections on the application of ChatGPT in remote banking. *Agricultural Bank Journal*, (05), 49–52. (In Chinese)
- [9] Ye, C. Y., & Zeng, D. D. (2023). Analysis of the application of ChatGPT and RPA technology. *Digital Communication World*, (09), 167–169+172. (In Chinese)
- [10] Qiu, D. Y., & Lan, Y. (2023). Opportunities, challenges, and issues brought by ChatGPT to the financial industry. *Southwest Finance*, (06), 18–29. (In Chinese)
- [11] Zeng, X. (2023). Discussion on the application of ChatGPT in the financial industry. *Financial Technology Era*, 31(05), 19–22. (In Chinese)
- [12] Chen, Y. W. (2023). Beyond ChatGPT: Opportunities, risks, and challenges of generative AI. *Journal of Shandong University (Philosophy and Social Sciences Edition)*, (03), 127–143. (In Chinese)
- [13] Wang, Y. M., Wang, D., Liang, W. Y., & Liu, C. C. (2023). "Aladdin's lamp" or "Pandora's box": Potential and risks of ChatGPT in educational applications. *Research on Modern Distance Education*, 35(02), 48–56.
<https://doi.org/10.3969/j.issn.1009-5195.2023.02.006> (In Chinese)
- [14] Jiao, J. L. (2023). ChatGPT promoting the digital transformation of school education--What to learn and how to teach in the age of artificial intelligence. *China Distance Education*, 43(04), 16–23.
<https://doi.org/10.13541/j.cnki.chinade.20230226.001> (In Chinese)