A Conversational Application for Insomnia Treatment: Leveraging the ChatGLM-LoRA Model for Cognitive Behavioral Therapy

Yinda Chen[1*], Shuo Pan[1*], Yu Xia[1*], Keqi Ren[1*], Luying Zhang[5], Zefei Mo[6], Jiahe Chen[7], Meijia Zhang[8], Huanhuan Li[9], Jianwei Shuai[10], Qinghua Xia[11#], Rongwen Yu[12#]

***Abstract*—The aim of this study was to develop a mobile application for psychotherapy with insomnia patients using the ChatGLM-LoRA model, fine-tuned by Low-Rank Adaptation, and validated in a clinical trial.**

The dataset used to train the model was a collection of 764 dialogues related to sleep disorders. The corpus was randomly divided into three subsets: training, validation, and test sets. The hyperparameters used in this study to train the model were 450 epochs, betas ranging from 0.9 to 0.95, weight decay rate 5e-4, maximum learning rate 1e-5, and AdamW optimizer. Based on the test results of the above hyperparameters, the four metrics of BLEU-4, ROUGE-1, ROUGE-2, and ROUGE-L of the model reached 0.0340, 0.0451, and 0.0163; 0.2773, 0.3075, and 0.1986; 0.0592, 0.0735, and 0.0261; 0.2112, 0.2336, and 0.1500 for the training, validation, and test sets.

These results indicate the technical feasibility and potential clinical utility of using an advanced language model-based application for psychotherapeutic intervention in insomnia.

***Index Terms*—large language model (LLM), sleep disorders, Low-Rank Adaptation (LoRA), application (APP)**

I. Introduction

A. Research-Background

Sleep disorders encompass a wide range of conditions that disrupt the normal pattern and quality of sleep, ranging from excessive sleepiness to insomnia, as well as various abnormal behaviors manifested during sleep. Among these, insomnia represents a significant segment of sleep disorders, with statis tics indicating that approximately 30% of adults are affected by insomnia annually [1]. The consequences of insomnia are extensive, not only leading to immediate problems such as daytime sleepiness and decreased concentration, but also posing long-term risks such as increased susceptibility to cardiovascular disease, depression, and anxiety [2]. Walker (2017) emphasized that the effects of insomnia extend beyond physical disturbances and significantly affect mental health. The U.S. Healthy People 2020 program has made ensuring adequate sleep duration one of its priority goals, highlighting the importance of adequate sleep to public health [3]. Given the urgency of this challenge, there is an urgent need to identify effective treatments for insomnia.

Current therapeutic approaches to insomnia are divided into pharmacologic and nonpharmacologic treatments. Pharmacologic treatments, especially benzodiazepines, are widely used; however, patients may experience adverse effects ranging from dizziness and drowsiness to severe cognitive impairment in the short term. Long-term use may lead to increased drug resistance and dependence. Non-pharmacological treatments are categorized into cognitive therapy, stimulus control, sleep restriction practices, sleep hygiene education, and relaxation training. Among these interventions, cognitive behavioral therapy for insomnia (CBT-I) is the primary method. CBT-I, a psychological approach, treats the disorder by intervening in an individual's cognition and behavior. It improves sleep quality without the side effects associated with medications, and its positive effects are highly durable [4]. In 2016, the American College of Physicians officially endorsed CBT-I as the foremost treatment strategy for managing insomnia [5]. Subsequently, in 2019, a thorough analysis combining data from 30 CBT-I studies highlighted its effectiveness in reducing the time needed to fall asleep, decreasing nocturnal awakenings, and enhancing sleep quality among individuals with insomnia. Notably, these benefits were sustained over a long period [6]. A randomized controlled trial conducted in 2023 further validated that CBT-I significantly improved several key

This work is supported by Ministry of Science and Technology of the People's Republic of China (STI2030-Major Projects2021ZD0201900), National Natural Science Foundation of China (Grant No. 12090052).

*Yinda Chen, Shuo Pan,Yu Xia and Keqi Ren contributed equally to this work.

#Qinghua Xia and Rongwen Yu contributed equally to this work.

[1]Yinda Chen is with the institute for School of Electrical and Information Engineering, Quzhou University, Zhejiang Province, China, 324000.(yd-chen@zeroacademy.net)

[2]Shuo Pan, Yu Xia, Keqi Ren, Luying Zhang, Jiahe Chen and Meijia Zhang are with the institute for Wenzhou Medical University, Wenzhou, China, 325001.(span@zeroacademy.net, yxia@zeroacademy.net, kqren@zeroacademy.net, lyzhang@zeroacademy.net, jhchen@zeroacademy.net and mjzhang@zeroacademy.net)

[3]Zefei Mo is with the institute for School of Ophthalmology and Optometry, Eye Hospital, Wenzhou Medical University, Wenzhou, China, 325001 (zfmo@zeroacademy.net)

[4]Huanhuan Li is with the institute for Zhejiang University, Ningbo, China,310058 (22260474@zju.edu.cn)

[5]Jianwei Shuai and Rongwen Yu are with the institute for Wenzhou Institute,University of Chinese Academy of Sciences, Wenzhou, China, 325001 (mailto:shuaijw@wiucas.ac.cn and rwyu@ucas.ac.cn)

[6]Qinghua Xia is with the institute for Ningbo Innovation Center,Zhejiang University, Ningbo, China,310058 (xiaqinghua@zju.edu.cn)

indicators, including the severity of insomnia, sleep onset latency, wakefulness after sleep onset, frequency of early morning awakenings, and overall sleep efficiency [7]. This marked a significant shift in the approach to treating insomnia—moving away from predominantly pharmacological treatments towards embracing safer and more effective non-pharmacological interventions such as psychological counseling.

However, despite the important role and cost-effectiveness of CBT-I in the treatment of insomnia, its widespread use is limited by the varying socio-economic development of countries and the relative scarcity of resources for psychotherapy [8]. In this context, the emergence of emerging artificial intelligence (AI) technology has provided an innovative solution. The application of AI in the field of mental health has yielded numerous positive outcomes, including the development of novel treatment strategies, outreach to patient populations that are traditionally difficult to engage, improved patient response rates, and the saving of valuable time for healthcare professionals [9]. AI technology allows us to effortlessly overcome challenges related to time constraints, geographic barriers, and scarcity of psychologist resources. It also provides a means to reduce or even eliminate the significant costs associated with traditional counseling sessions. This approach is not only safer and more effective, but also more affordable, positioning it as a superior alternative for the treatment of insomnia.

B. Content of the Study

This study focuses on the design and development of a conversational application using the ChatGLM model. The goal is to develop a system capable of addressing various clinical queries related to sleep and providing relevant advice and guidance to both patients and healthcare professionals. By utilizing extensive online data resources on CBT-I, the system strives to perform in-depth analyses, thereby providing highly individualized sleep management programs and recommendations that meet the diverse needs of users.

In our research, we use the ChatGLM large language model as a fundamental benchmark, which integrates self-attention mechanisms and multi-attention technologies, enabling it to adeptly identify long-range dependencies and intricate contextual nuances within texts. This capability is achieved by generating predictive results from the output layer after feature extraction via the deep network architecture of the Transformer encoding layer. To enhance model generalization, we implement the Low-Rank Adaptation (LoRA) technique and select AdamW as the optimizer for fine-tuning model parameters. Model performance is automatically evaluated using the c-eval database, with accuracy rate serving as a quantitative metric to measure improvements in model behavior before and after enhancements. This methodological approach allows for a professional and objective evaluation of the effectiveness of the ChatGLM model in real-world applications after optimization with LoRA technology. To enhance the practical utility, we integrate the ChatGLM-LoRA model, trained over 450 epochs, into our core framework using Flask for backend development. This facilitates the creation of a sleep support application. To validate its clinical relevance, we enlist 16 volunteers to provide feedback over a one-week period, allowing us to comprehensively assess the practical significance of the model in clinical settings.

II. Literature Studies

The beginning of the 21st century has witnessed a significant surge in artificial intelligence (AI) technology, manifesting a pronounced potential in the delivery of targeted psychological interventions, particularly in the area of sleep counseling. A growing body of scientific work underscores the burgeoning application of large language model (LLM)-based AI dialog systems in achieving therapeutic goals in mental health care [10]. For example, Vaswani et al. (2017), in their seminal paper "Attention Is All You Need," elucidated how the Transformer model facilitates counseling services related to sleep disorders. At the same time, Bojic et al. introduced a hybrid human-AI health training paradigm that integrates a sleep-focused Q&A dataset [11].

A growing body of empirical evidence supports the efficacy of network-delivered cognitive behavioral therapy (CBT)-based interventions for insomnia [12]. In particular, CBT-oriented mental health chatbots, such as those developed by Woebot, Wysa, and Tess, have demonstrated considerable success in improving both the mental and physical well-being of users [10]. In addition, there have been notable advances in the incorporation of AI agents-including chatbots-into digital health interventions. These developments have been instrumental in managing symptoms and promoting health-promoting behaviors [13].

Additionally, in the field of sleep medicine, the application of big data technologies has been recognized for its ability to effectively monitor, analyze, and predict problems associated with sleep disorders [3]. This includes the use of machine learning algorithms and cognitive strategies, along with existing knowledge bases, to identify abnormal sleep behaviors [14]. Erica Corda and colleagues pioneered the development of a predictive sleep system, accompanied by an APP. This system combined machine learning algorithms and LLMs, and demonstrated its effectiveness through empirical research using real-world data. It also investigated the utility of extensive linguistic modeling in improving sleep quality [15]. In addition, several innovative models were introduced, such as psycholinguistic-based models (e.g., LIWC and Empath), bidirectional encoder representations of transformers (e.g., BERT), and Big 5 personality-based models. These frameworks supported the construction of comprehensive approaches to the analysis and prediction of insomnia [16].

Despite these advances, the use of large language models in assistive systems faces several challenges. While the use of LLMs facilitates the creation of datasets to some extent, the objective evaluation of the results generated by these models remains problematic in practice [17]. Furthermore, despite the contributions of AI technology in promoting sleep health, it still falls short of clinical psychologists in providing personalized care, tailored programs, and diverse treatment strategies when compared to traditional psychotherapy services. Therefore, it is imperative to enhance AI models with broader databases and more sophisticated algorithms. At the same time, there is an urgent need to improve the adaptability of these models to real-world conditions to better meet the individual needs of users. Looking forward, it is imperative to explore more nuanced and comprehensive methodologies to increase the application value and efficiency of AI in sleep medicine.

III. Research Methods

A. Implementation Strategy

The methodology employed in this study comprises four key phases: data collation, model building, application development, and application utilization. These phases, delineated in Fig. 1, collectively form the backbone of our investigative approach.

B. Equipment Environment

Our research infrastructure is divided into software and hardware components. In terms of software, we used Android Studio, a software used by researchers for coding to create applications. The use of Flask-based Python as a server backend [18] facilitated seamless data transfer between local machines and servers via okhttputils. Secure Shell (SSH) protocols were instrumental in configuring and managing server settings, ensuring a secure and efficient flow of data. In particular, Google's Material Design framework was adopted for secondary development, enhancing the aesthetics and functionality of various interface components.

For our hardware configuration, we utilized an A100-PCIE-40GB system, which comprises a Xeon Gold 6248R CPU, 72GB of RAM, 40GB of graphics memory, 50GB of storage, and an operating system that conforms to the standards required for developing chatbot applications.

C. Recruiting Volunteers to Use the Application

To determine the effectiveness and usability of our application, we conducted a rigorous clinical trial with volunteer participants. We devised a multi-faceted recruitment strategy and advertised extensively on the Wenzhou Medical University campus to attract a diverse pool of potential candidates. We clearly explained the content and objectives of the research project to them and confirmed their willingness to participate. For the initial recruitment, which yielded 32 interested volunteers, we recorded their information and conducted further qualification checks. The volunteer screening criteria are detailed in Fig. 2. Through this process of screening, we screened a total of 16 volunteers who met the inclusion criteria and were subsequently enrolled to participate in the clinical trial.

IV. Results and Discussion

A. Data Collation

Between November 28 and December 14, 2023, our study collected a substantial dataset comprising 21,924 records through five distinct avenues: instances, media, literature, guidelines, and databases. To ensure the integrity and relevance of our data, we undertook a meticulous cleaning process spearheaded by three experienced quality controllers. Each controller boasts broad expertise in mental health counseling. Initially, we eliminated duplicates and linguistic inaccuracies. Subsequently, we focused on screening data for key terms such as “sleep”, “dream”, “evening”, “night”, “bed” and related expressions. This step aimed to sift out conversational data irrelevant to our study's objectives, ensuring that the textual content was pertinent to the context of potential sleep disorders. We further conducted dialog integrity screening. Finally, through manual inspection on a case-by-case basis, we excluded dialogue data that contradicted universal core values. This rigorous curation process resulted in the selection of 764 dialogues that aligned with our study criteria, as depicted in Fig. 3. All dialogue data utilized in this research obtained ethical clearance from the Ethics Review Committee of the First Affiliated Hospital of Wenzhou Medical University. This approval ensures adherence to established ethical standards.

University, ensuring compliance with established ethical standards. To facilitate systematic analysis, the final dataset of 764 dialogues was randomly divided into three subsets: 80% for the training set (611 dialogues), 10% for the validation set (77 dialogues), and 10% for the test set (76 dialogues). The allocation of conversations within each subset was carefully designed to ensure randomness and balance, contributing to the robustness of our analyses. The resulting language model is accessible via an application programming interface (API).

B. Model Building

1) Model Selection: In this investigation, we used the ChatGLM large language model, which is basically built around the GLMBlock. This central component uses the sophisticated Transformer architecture, which integrates a self-attention mechanism along with a multi-head attention strategy. It also includes critical techniques such as Add & Layer Norm and Gated Linear Units (GLU). In addition, we have extended the GLMBlock with Layer Norm and Dropout Layers to mitigate the problem of gradient vanishing, reduce overfitting, and improve the model's ability to generalize. These modifications provide the model with an enhanced ability to capture long-range dependencies and intricate contextual information within textual data [19]. In the manuscript, word embeddings coupled with positional encoding are used to preprocess the input data. The deep network architecture of the Transformer's coding layers then performs feature extraction. The output layer is then responsible for generating predictive results. Known for its broad applicability in natural language processing, ChatGLM delivers superior speech understanding and generation capabilities due to its sophisticated architecture and advanced optimization algorithms. Compared to models of similar size, ChatGLM provides optimal performance while ensuring low resource consumption, achieved through refined training methods and strategic algorithmic optimization [20].

2) Model Tuning and Training: Due to the complex nature of medical terminology and the wide variety of textual material, traditional fine-tuning methods often lead to an overfitting scenario within the model, in addition to requiring significant computational resources. To overcome these challenges, we have incorporated the LoRA technique into our model tuning process [21]. The LoRA approach efficiently tunes model parameters by integrating two trainable matrices characterized by low-rank decomposition into the model parameter space. This is achieved without significant computational cost [22], [23].

In the linear layer configuration, the weight matrix is represented as $W_0 \in R^{d \times k}$ where k is the input dimension and d is the output dimension. LoRA introduces two trainable matrices with low-rank decompositions, denoted as $B \in R^{d \times r}, A \in R^{r \times k}$, where r is the predetermined rank. The forward propagation formula is modified to be:

$$h = Wh = W_0 x + \delta W x = W_0 x + BAx, B {\in}^{d \times r}, A {\in}^{r \times k} \tag{1}$$

After fine-tuning with LoRA, we obtained the corresponding fine-tuned checkpoints for subsequent testing phases. A salient feature of LoRA is its ability to reduce hardware resource consumption while maintaining training efficiency. This advantage allows for more feasible fine-tuning of larger models under equivalent memory conditions, bringing greater flexibility and efficiency to research efforts.Fig. 4shows the network architecture diagram of our model.

3) Model Evaluation and Preservation: In this research, four metrics (BLEU-4 and ROUGE-1, 2, L) were used to evaluate the performance of the model and to quantify its quality using precision and recall rates. Specifically, BLEU-4 is used as a critical metric for assessing the quality of machine translation. This metric measures the quality of the translation by calculating the n-gram concordance between the text generated by the model and a reference standard. A high BLEU score indicates a high degree of n-gram overlap, which is generally indicative of superior translation fidelity. The ROUGE-1, 2, L suite of metrics is primarily concerned with assessing the coverage and retention between content produced by an automated summarization or machine translation system and its reference material. The ROUGE-N metric (where N is 1 or 2) primarily assesses the overlap of N contiguous units, such as words, between texts. Conversely, ROUGE-L is designed to assess the length of the longest common subsequence, thus serving as an indicator of structural congruence within sentences.

After 450 training cycles, the GLM LoRA enhanced model was evaluated using the training, validation and test sets.Fig. 5 illustrates the evolution of the performance metrics over the course of training for the three datasets using line graphs, using the BLEU-4 and ROUGE-1, ROUGE-2 and ROUGE-L metrics. After 450 training rounds, the metrics of our model BLEU-4, ROUGE-1, ROUGE-2, ROUGE-L reach 0.0340, 0.0451, 0.0163; 0.2773, 0.3075, 0.1986; 0.0592, 0.0735, 0.0261; 0.2112, 0.2336, 0.1500 in the training, validation and test sets, respectively. Finally, we selected the iterative version of the BLEU metric that showed the best results on the validation set and saved its corresponding parameters as our final adopted model.

After the ChatGLM model was optimized to extend its task-specific capabilities, it was evaluated using C-Eval, a robust and impressive benchmark for evaluation. The C-Eval test provides the change in average accuracy (Accuracy) across models of the same size, with Accuracy decreasing from 47.3684 to 36.8421 after 450 epochs of training. The Application Utilization section in Fig. 1 shows how well our model performs.

C. Application Development

In order to provide patients with better quality support and assistance, and to help doctors make faster diagnostic decisions - thus reducing waiting time for patients - we have developed and designed an application equipped with a question-and-answer sleep support system based on conversational AI technology. By interacting with the system via mobile, patients can overcome geographical limitations to receive real-time professional advice and guidance on sleep disorders and receive personalized treatment plans. The Android-based sleep support application described in this manuscript uses the Flask framework to orchestrate Python-based web services. Central to its operation is the use of the sophisticated ChatGLM-LoRA model, which serves as the core algorithm. Interaction with the system is facilitated by a RESTful API interface using the POST method, which optimizes request handling and data responsiveness. The backend architecture uses JSON for data exchange, promoting both flexibility and efficiency in data interactions. On the front-end, development is anchored in Android Studio, with Gradle managing dependent packages to ensure a smooth and streamlined development workflow. To enhance user privacy and security, the application uses JSON files to store user data on the server side, complemented by a SQLite database to strengthen data protection on the local front. In addition, to enrich the user interaction experience, the application interface uses a ListView to display the dialog list. Strict adherence to Android security best practice guidelines was a cornerstone throughout the development process, ensuring the application's robust security and reliability. Extensive Android device compatibility testing was methodically performed to ensure the application's compatibility across a wide range of devices and to maintain a consistent user experience.

D. Application Utilization

Sixteen volunteers were recruited for this study, each of whom underwent a one-week intervention of conversational counseling via an app incorporating an optimized ChatGLM model. The study aimed to provide participants with sleep advice and anxiety reduction.At the end of the experiment, the volunteers completed the questionnaire of the software usage research, and we further statistically analyzed the collected questionnaire information. According to the questionnaires, more than 50%(8 volunteers) of the volunteers thought that the information provided by the app effectively improved their sleep, and 75%(12 volunteers) of the volunteers were willing to continue to use the app and recommend it to others. More than 80% of the volunteers think that our app is well designed and smooth to use, and the total number of clicks on the app has exceeded 810.

V. Conclusions and Suggestionss

Based on the analysis and design results, we have come to several pertinent conclusions:

A) A sleep quiz model was developed using the ChatGLM large language model. In addition, a companion application to support sleep quizzes was designed and implemented.

B) Using the AdamW optimizer, a maximum learning rate of 1e-5 was achieved. Betas were kept in the range of 0.9 to 0.95, and a weight decay rate of 5e-4 was set. This configuration showed stable performance over 450 training iterations, with loss metrics fluctuating between 1.4 and 1.6. Furthermore, c-eval testing showed an accuracy of 36.8421%. However, test results suggest that while the model's adaptability to specific tasks improved, its overall accuracy showed a declining trend.

This observation may indicate that fine-tuning aimed at improving task-specific performance potentially compromises the model's ability to generalize to a broader range of tasks. This underscores the critical importance of careful metric selection and sensitivity considerations in the development of scoring systems-a sensitive and comprehensive scoring mechanism is critical for accurately assessing the impact of fine-tuning on model effectiveness. Consequently, this underscores the need for a balanced approach to model fine-tuning and evaluation that aims to improve task-specific performance while maintaining generalizability. Thus, the use of a variety of evaluation techniques, including complex task test sets such as C-Eval, is advisable to ensure both specialized task efficiency and broader scenario adaptability.

C) User experience feedback indicates that more than 50% of the volunteers felt that using the information provided by the app was effective in improving their sleep, showing that the model performed satisfactorily in real-world application contexts.

Future research directions include:

a) Enriching the training corpus with more diverse databases to strengthen the model's generalization capabilities across different scenarios.

b) Integrating additional deep learning large language models and refining fine-tuning methods to improve overall model performance - effectively addressing a wider range of individual needs.

c) Engage in continuous iterative optimization based on user feedback to dynamically adjust strategies in response to evolving needs.

References

[1] C. M. Morin and D. C. Jarrin, "Epidemiology of insomnia: prevalence, course, risk factors, and public health burden," *Sleep medicine clinics*, vol. 17, no. 2, pp. 173–191, 2022.

[2] D. Riemann, F. Benz, R. J. Dressle, C. A. Espie, A. F. Johann, T. F. Blanken, J. Leerssen, R. Wassing, A. L. Henry, S. D. Kyle *et al.*, "Insomnia disorder: State of the science and challenges for the future," *Journal of sleep research*, vol. 31, no. 4, p. e13604, 2022.

[3] N. L. Bragazzi, O. Guglielmi, and S. Garbarino, "Sleepomics: how big data can revolutionize sleep science," *International journal of environmental research and public health*, vol. 16, no. 2, p. 291, 2019.

[4] Y. Zhu, C. Stephenson, E. Moghimi, J. Jagayat, N. Nikjoo, A. Kumar, A. Shirazi, C. Patel, M. Omrani, and N. Alavi, "Investigating the effectiveness of electronically delivered cognitive behavioural therapy (e-cbti) compared to pharmaceutical interventions in treating insomnia: Protocol for a randomized controlled trial," *Plos one*, vol. 18, no. 5, p. e0285757, 2023.

[5] A. Van Straten, T. van der Zweerde, A. Kleiboer, P. Cuijpers, C. M. Morin, and J. Lancee, "Cognitive and behavioral therapies in the treatment of insomnia: a meta-analysis," *Sleep medicine reviews*, vol. 38, pp. 3–16, 2018.

[6] T. van der Zweerde, L. Bisdounis, S. D. Kyle, J. Lancee, and A. van Straten, "Cognitive behavioral therapy for insomnia: a meta-analysis of long-term effects in controlled studies," *Sleep medicine reviews*, vol. 48, p. 101208, 2019.

[7] Y. Takano, R. Ibata, N. Machida, A. Ubara, and I. Okajima, "Effect of cognitive behavioral therapy for insomnia in workers: A systematic review and meta-analysis of randomized controlled trials," *Sleep Medicine Reviews*, p. 101839, 2023.

[8] A. N. Natsky, A. Vakulin, C. L. Chai-Coetzer, L. Lack, R. McEvoy, N. Lovato, A. Sweetman, C. J. Gordon, R. J. Adams, and B. Kaambwa, "Economic evaluation of cognitive behavioural therapy for insomnia (cbt-i) for improving health outcomes in adult populations: a systematic review," *Sleep Medicine Reviews*, vol. 54, p. 101351, 2020.

[9] A. Fiske, P. Henningsen, and A. Buyx, "Your robot therapist will see you now: ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy," *Journal of medical Internet research*, vol. 21, no. 5, p. e13216, 2019.

[10] J. Grodniewicz and M. Hohol, "Waiting for a digital therapist: three challenges on the path to psychotherapy delivered by artificial intelligence," *Frontiers in Psychiatry*, vol. 14, p. 1190084, 2023.

[11] I. Bojic, Q. C. Ong, M. Thakkar, E. Kamran, I. Y. Le Shua, J. R. E. Pang, J. Chen, V. Nayak, S. Joty, and J. Car, "Sleepqa: A health coaching dataset on sleep for extractive question answering," in *Machine Learning for Health*. PMLR, 2022, pp. 199–217.

[12] K. M. Shaffer, E. A. Finkelstein, F. Camacho, K. S. Ingersoll, F. Thorndike, and L. M. Ritterband, “Effects of an internet-based cognitive behavioral therapy for insomnia program on work productivity: a secondary analysis,” *Annals of Behavioral Medicine*, vol. 55, no. 6, pp. 592–599, 2021.

[13] H. Shim, “Development of conversational ai for sleep coaching programme,” in *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Student Research Workshop*, 2021, pp. 121–128.

[14] A. Coronato and G. Paragliola, “Towards a cognitive system for the identification of sleep disorders,” in *Intelligent Interactive Multimedia Systems and Services 2017 10*. Springer, 2018, pp. 91–98.

[15] E. Corda, S. M. Massa, and D. Riboni, “Context-aware behavioral tips to improve sleep quality via machine learning and large language models,” *Future Internet*, vol. 16, no. 2, p. 46, 2024.

[16] A. S. Sakib, M. S. H. Mukta, F. R. Huda, A. N. Islam, T. Islam, and M. E. Ali, “Identifying insomnia from social media posts: psycholinguistic analyses of user tweets,” *Journal of medical Internet research*, vol. 23, no. 12, p. e27613, 2021.

[17] S. Trott, “Can large language models help augment english psycholinguistic datasets?” *Behavior Research Methods*, pp. 1–19, 2024.

[18] Z. Yi, J. Ouyang, Y. Liu, T. Liao, Z. Xu, and Y. Shen, “A survey on recent advances in llm-based multi-turn dialogue systems,” *arXiv preprint arXiv:2402.18013*, 2024.

[19] K. Hulliyah, F. Rayyan, and N. S. A. A. Bakar, “Development of a chatbot for the online application telegram chat with an approach to the emotion classification text using the indobert-lite method,” in *2022 4th International Conference on Cybernetics and Intelligent System (ICORIS)*. IEEE, 2022, pp. 1–4.

[20] C. Liu, K. Sun, Q. Zhou, Y. Duan, J. Shu, H. Kan, Z. Gu, and J. Hu, “Cpmi-chatglm: parameter-efficient fine-tuning chatglm with chinese patent medicine instructions,” *Scientific Reports*, vol. 14, no. 1, p. 6403, 2024.

[21] V.-T. Doan, Q.-T. Truong, D.-V. Nguyen, V.-T. Nguyen, and T.-N. Nguyen Luu, “Efficient finetuning large language models for vietnamese chatbot,” *arXiv e-prints*, pp. arXiv–2309, 2023.

[22] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, “Lora: Low-rank adaptation of large language models,” *arXiv preprint arXiv:2106.09685*, 2021.

[23] A. Chavan, Z. Liu, D. Gupta, E. Xing, and Z. Shen, “One-for-all: Generalized lora for parameter-efficient fine-tuning,” *arXiv preprint arXiv:2306.07967*, 2023.

FIGURES

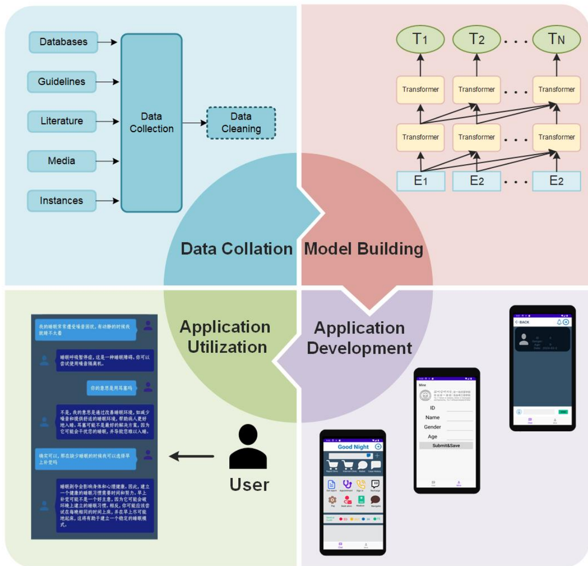

Fig. 1. Experimental method demonstration diagram.

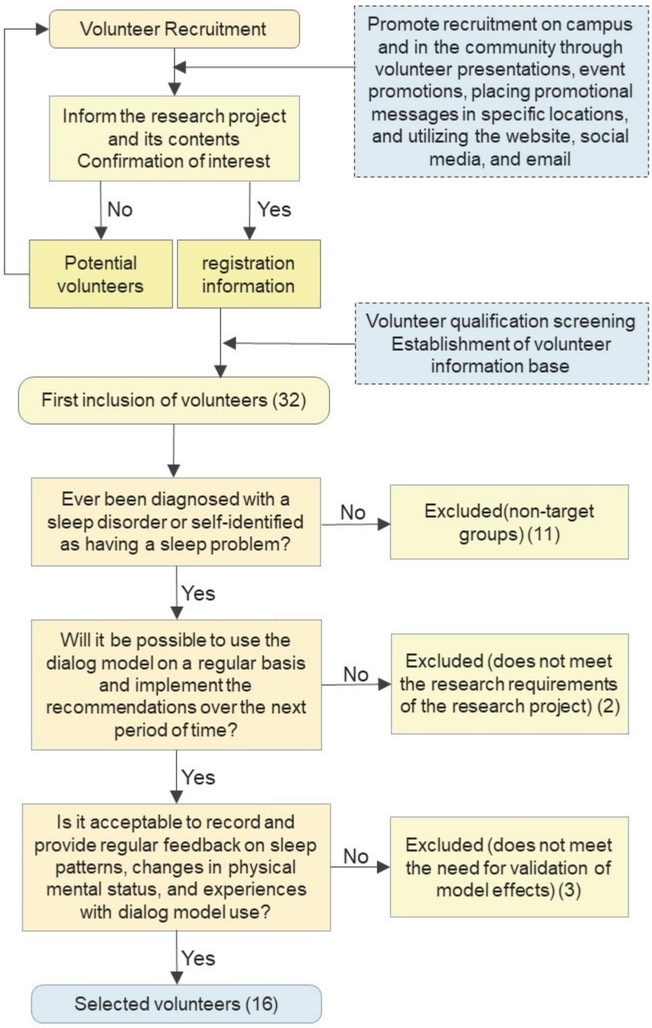

Fig. 2. Volunteer Screening Chart

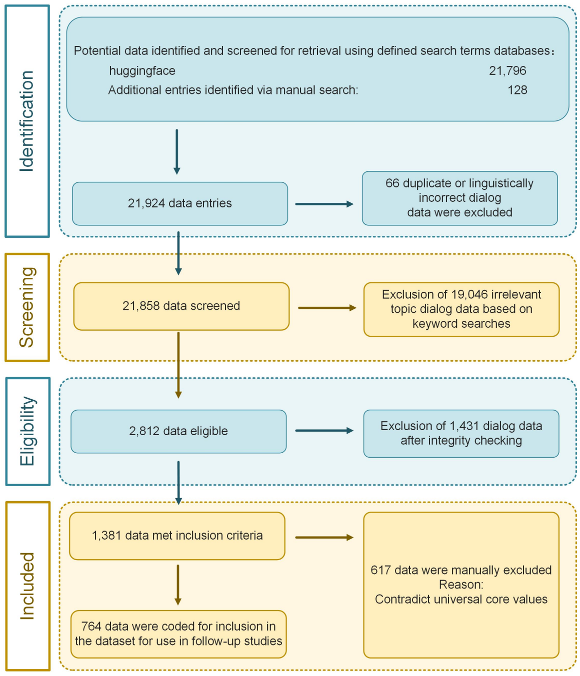

Fig. 3. Data entry and exit group diagram. The figure shows the entire process from data collection to data screening to final data determination.

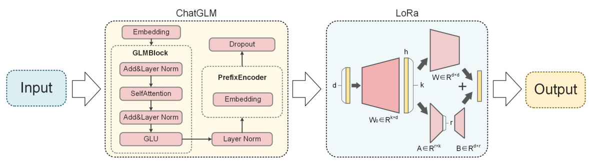

Fig. 4. Architecture Diagram of the ChatGLM-based Large-Scale Conversational Language Model

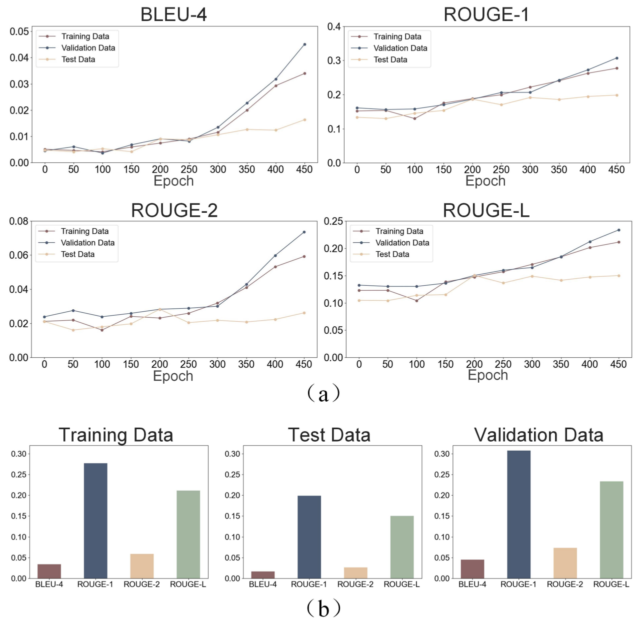

Fig. 5. Model Performance Diagram :(a) BLEU-4, ROUGE-1, ROUGE-2, and ROUGE-L metrics change folds for training 450epoch, training set, validation set, and test set(b) Comparison of BLEU-4, ROUGE-1, ROUGE-2, ROUGE-L metrics for final training set, validation set, and test set