Enhancing Nursing and Elderly Care with Large Language Models: A Framework for AI-Driven Patient Monitoring and Interaction

Anonymous COLING 2025 submission

Abstract

This paper explores the application of large language models (LLMs) in nursing and elderly care, focusing on AI-driven patient monitoring and interaction. We introduce a novel Chinese nursing dataset and implement incremental pre-training (IPT) and supervised fine-tuning (SFT) techniques to enhance LLM performance in specialized tasks. Using LangChain, we develop a dynamic nursing assistant capable of real-time care and personalized interventions. Experimental results demonstrate significant improvements, paving the way for AI-driven solutions to meet the growing demands of healthcare in aging populations.

1 Introduction

The rapid advancement of large language models (LLMs) has opened new avenues for healthcare applications. While LLMs have demonstrated impressive capabilities in generating human-doctor-like clinical decisions and integration into healthcare (Thirunavukarasu et al., 2023; Tan et al., 2024; Ullah et al., 2024; Li et al., 2023), its expertise in nursing remains in its nascent stages.

On the one hand, nursing scenarios are more complex than other clinical decision cases, such as medication prescription or diagnostic imaging, as they involve continuous monitoring, real-time decision-making, and patient interaction, requiring models that can handle a wider array of multimodal inputs and adapt dynamically to evolving patient conditions (Carayon and Gurses, 2008). On the other hand, nursing tasks often involve high levels of direct patient interaction, demanding models that can process complex multimodal inputs—such as voice, text, and even visual cues—in real time. China has experienced a significant increase in its aging population. By 2022, individuals aged 60 and above accounted for 19.8% of the population (Global Times, 2023), a figure projected to rise to 28% by 2040 (Peng, 2023). This demographic shift is expected to place considerable pressure on the country's healthcare system, particularly in meeting the demand for nursing care. Despite this growing need, the supply of skilled nursing services remains inadequate. There is a noticeable gap between the expertise required to care for the elderly and the qualifications of the current healthcare workforce. A 2023 investigation revealed that only 7.18% of workers in China's elderly care industry hold a bachelor' s degree or higher, highlighting the urgent need for more qualified personnel (Zhang and Zhang, 2023).

Our work seeks to address this disparity by developing AI-driven Nursing and Elderly-Care solutions tailored to the specific needs of the nursing profession, leveraging cutting-edge large language model (LLM) techniques. We focus on creating LLMs that support patient monitoring, personalized care, and facilitate effective communication between healthcare providers and patients. Additionally, we are exploring the development of a Langchain agent application based on this specialized model, alongside its potential for multimodal processing.

In this paper, we make the following contributions to both the NLP community and the Nursing and Elderly Care industry:

- We pioneered the application of large language models in nursing and elderly care, proposing a SOTA model and gathering fine-tuning expertise specific to these fields.
- We developed the first multilayer Chinese nursing dataset for elderly care and

demonstrate its effectiveness through ablation studies. We also establish a benchmark test set to evaluate fundamental nursing knowledge and skills.

- We investigate the use of nursing robots powered by our LLM, evaluating their performance in essential nursing tasks and exploring their potential to incorporate visual processing in care environments.

2 Related Work

2.1 Harnessing LLMs for Nursing Applications

Studies have highlighted the transformative role of Large Language Models (LLMs) in healthcare, including applications in clinical decision-making, patient care, and medical education. Comprehensive surveys discuss the development and deployment of LLMs across various medical tasks, focusing on their potential for improving diagnostic accuracy and streamlining medical workflows (Zhou et al., 2023; Nazi and Peng, 2024). (Zhou et al., 2023) also highlights the performance of models like GPT-4 and MedPaLM across ten biomedical natural language processing tasks, demonstrating their generalization ability to outperform traditional models in various discriminative and generative tasks. However, the existing body of work mainly focuses on the general medical applications of LLMs, while LLMs specifically designed for nursing applications are left under-explored.

Nursing environments present a unique set of challenges, such as time-sensitive decision-making, handling diverse patient populations, and managing high-stress situations. Current general medical LLMs may not be fully equipped to address these demands, emphasizing the need for more focused research on LLMs designed specifically for nursing applications.

The current related work in nursing primarily focuses on theoretical exploration and future possibilities, rather than practical implementation. (Xiong et al., 2023) validates the combination of LLMs with local knowledge bases for intelligent nursing decision-making, highlighting the importance of contextual adaptation. However, their model is developed solely on the text modality, lacking integration with other crucial data sources such as audio and visual inputs. Other work (Perspect, 2023; Woo et al., 2024) discusses the implications of LLMs in healthcare education, noting both their potential and the need for cautious implementation. These studies collectively underscore the importance of contextual, safe, and practical integration of LLMs in nursing.

The most closely related work to ours is LlamaCare (Li et al., 2024), a large language model that utilizes instruction-based tuning to integrate diverse clinical data, improving its ability to generate discharge summaries and predict outcomes like mortality and hospital stays. LlamaCare surpasses existing LLM benchmarks in producing accurate and coherent clinical texts, demonstrating its potential for broader clinical use. However, its focus spans a wide range of healthcare domains, with less emphasis on the foundational knowledge specific to nursing.

2.2 Nursing Datasets for LLMs

Developing nursing-specific datasets is essential for improving LLMs in healthcare, but such datasets are limited, restricting their application in specialized fields like nursing. While the MIMIC-III database (Johnson et al., 2016) offers structured data, it lacks alignment with the unstructured text needed for LLMs.

Wang et al. (Wang et al., 2023b) introduced MedNgage, a dataset focused on patient-nurse conversations, annotated to distinguish between socio-affective and cognitive engagement. Fine-tuning transformer models on this dataset enhances AI-driven predictions in patient care.

Xiong et al. (Xiong et al., 2023) developed a dataset that integrates LLMs with local knowledge bases for decision-making in nursing, but it primarily addresses textual data, lacking the multimodal inputs (e.g., audio, visual) essential for real-time patient interactions.

3 Method

3.1 Model Architecture

Our method builds upon cutting-edge large language models (LLMs) by applying supervised fine-tuning (SFT) to adapt these mod-

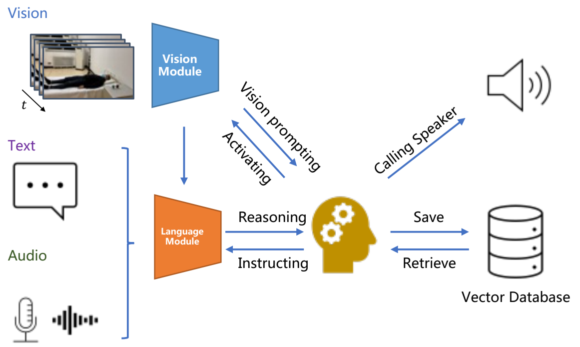

Figure 1: An overview for the multimodal agent framework.

els specifically for nursing and elderly care tasks. We primarily tested two advanced models: GLM4 (GLM et al., 2024) and LLaMA 3.1 (Vavekanand and Sam, 2024), both of which represent the state-of-the-art in LLM development, and can be integrated with multimodal ability easily via projection and fine-tuning (Wang et al., 2023a; Liu et al., 2024, 2023a,b).

3.2 Dataset

We developed a specialized dataset named "NursingPiles", designed to comprehensively cover various sources and levels of professional knowledge in nursing and elderly care. This dataset is built from multiple sources, including textbooks, manuals, legal documents, and research papers, synthesizing data into question-answer (QA) pairs. To mitigate catastrophic forgetting (Zhai et al., 2023), which can occur during model fine-tuning, we introduced open-source datasets as part of a data-mixing strategy. This approach helps maintain the model's original dialogue capabilities while fine-tuning it for specialized tasks in nursing care.

3.3 Training Protocol

For the model training, we utilized the Parameter-Efficient Fine-Tuning (PEFT) package along with an Incremental Pre-training (IPT) process to further optimize the model' s performance. The training was conducted on 8 × NVIDIA A100-80GB GPUs, with a total training time of approximately 72 hours for fine-tuning, while the IPT stage took an additional 30 hours. The parameter settings for both stages are presented in Appendix A Table 5.

3.4 LangChain Prompting

In this design, we present a modular system for a dynamic nursing assistant, capable of handling the full lifecycle of patient care, including real-time data collection, personalized care plan generation, and continuous monitoring. The system integrates IoT devices for health data collection, AI-based diagnostics, and personalized care recommendations through LangChain. Critical to the design is the secure storage and management of patient information, utilizing AES encryption and key management services (KMS) to ensure data protection. Additionally, we employ OAuth and JWT for robust authentication, ensuring authorized access to encrypted data, and provide post-care follow-up with automated reminders and health education. This architecture allows for flexible, secure, and scalable patient care management. Figure 1 illustrates the pipeline, with the Appendix A providing core code and snippets for key processes.

3.5 Benchmark

We selected several authoritative exam questions, such as the "Three Basics and Three Stricts" exam questions (Zhang, 2020) and the postgraduate nursing exam questions (Li, 2019), as evaluation benchmarks. The entire set of questions includes two parts: multiple-choice questions and open-ended questions. For the multiple-choice questions, the "Three Basics and Three Stricts" test covers content from nine subjects, including basic theory (such as anatomy, physiology, and pathology), basic knowledge (including pharmacology, microbiology, and disease studies), and basic skills (such as nursing procedures, emergency techniques, and nursing operations). These subjects can objectively and comprehensively reflect the nursing knowledge and capabilities of the model (Wang, 2018). For this part of the questions, we use the P-R-F1 metrics to evaluate.

4 Experiments

4.1 Test Scores

We evaluated the performance of the models using Precision, Recall, F1-score, and Ac-

Data Format	Source	Utilization Method	Scale
Text in markdown format	Textbooks	IPT	2,777,526 Tokens
	Manuals, Industry Regulations	RAG	497,184 Tokens
Single-turn dialogues	SelfQA based on research papers	PEFT	17,580 pairs
	QA based on nursing safety and ethics from manuals, regulations	PEFT	5,000 pairs
	Medical open-source datasets	PEFT	5,000 pairs
Multi-turn dialogues	Generated nursing dialogues in simulated scenarios (GPT-4)	PEFT	1M dialogues
	Psychology and clinical dialogues generated by GPT-4o	PEFT	0.5M dialogues
Image-text pairs	Real-world photo collection	SFT	2,510 pairs

Table 1: Summary of data formats, sources, utilization methods, and scale. Abbreviations: IPT (Incremental Pretraining), RAG (Retrieval-Augmented Generation), PEFT (Parameter-Efficient Fine-Tuning), SFT (Supervised Fine-Tuning).

curacy. The results demonstrate that our models, which integrate both Incremental Pretraining (IPT) and Supervised Fine-Tuning (SFT), significantly outperform the baseline models. The GLM4-Chat 9B + IPT + SFT achieved the best performance with a Precision of 86.78%, Recall of 85.65%, F1-score of 86.21%, and Accuracy of 58.9%. These improvements highlight the importance of combining domain-specific pretraining with fine-tuning. For more details see Tbale 2.

Models	Precision	Recall	F1	Accuracy
LLaMA 3.1 8B Instruct	76.61	67.4	71.71	36.6
GLM4-Chat 9B	82.54	77.8	80.1	44.0
GPT-4o	86.62	84.02	85.3	56.84
Ours				
LLaMA + IPT + SFT	77.41	78.09	77.75	44.7
GLM4 + IPT + SFT	**86.78**	**85.65**	**86.21**	**58.9**

Table 2: Performance comparison between models, with highest score in bold.

4.2 Ablation Analysis

To assess the individual contributions of IPT and SFT, we conducted an ablation study by removing each component separately. The results show that removing either IPT or SFT results in a drop in performance across all metrics. For instance, without SFT, the LLaMA + IPT model saw a significant reduction in Recall (from 78.09% to 72.5%) and F1-score (from 77.75% to 74.69%). Similarly, removing IPT resulted in reduced performance for both models, particularly in Accuracy. This confirms that both components are crucial for optimal model performance in the nursing and elderly care domain. For more details see Tbale 3.

5 Conclusion

This paper presented an approach to apply large language models (LLMs) in nursing and elderly care by utilizing incremental pre-training (IPT) and supervised fine-tuning (SFT). We developed a Chinese nursing dataset, demonstrating its effectiveness through improved performance in specialized tasks. Additionally, we explored the use of LangChain for a dynamic nursing assistant, enabling real-time monitoring and personalized care. Our results highlight the potential of LLMs to address the growing demand for skilled nursing care.

Models	Precision	Recall	F1	Accuracy
Ours				
LLaMA + Instruct + IPT + SFT	77.41 (–)	78.09 (–)	77.75 (–)	44.7 (–)
GLM4 + IPT + SFT	86.78 (–)	85.65 (–)	86.21 (–)	58.9 (–)
Ablation (IPT only)				
LLaMA + IPT	77.00 (-0.41)	72.5 (-5.59)	74.69 (-3.06)	41.0 (-3.7)
GLM4 + IPT	85.50 (-1.28)	82.5 (-3.15)	84.0 (-2.21)	50.0 (-8.9)
Ablation (SFT only)				
LLaMA + SFT	76.90 (-0.51)	73.2 (-4.89)	74.98 (-2.77)	40.5 (-4.2)
GLM4 + SFT	86.00 (-0.78)	83.0 (-2.65)	84.48 (-1.73)	52.5 (-6.4)

Table 3: Performance comparison between models, with delta values shown in parentheses representing the difference between the full model (IPT + SFT) and the ablation variants.

6 Limitations

There are several concerns with respect to the limitations:

First, the model primarily focuses on text-based data, and further integration of audio and visual inputs is needed. Second, the dataset is largely Chinese-focused, limiting broader applicability across languages and cultures. Third, model responsiveness in real-time clinical settings remains a challenge. Last, ensuring patient privacy, consent, and minimizing bias in AI-driven care requires further consideration.

References

Pascale Carayon and Ayse P Gurses. 2008. Nursing workload and patient safety—a human factors engineering perspective. *Patient safety and quality: An evidence-based handbook for nurses.*

Team GLM, Aohan Zeng, Bin Xu, Bowen Wang, Chenhui Zhang, Da Yin, Diego Rojas, Guanyu Feng, Hanlin Zhao, Hanyu Lai, Hao Yu, Hongning Wang, Jiadai Sun, Jiajie Zhang, Jiale Cheng, Jiayi Gui, Jie Tang, Jing Zhang, Juanzi Li, Lei Zhao, Lindong Wu, Lucen Zhong, Mingdao Liu, Minlie Huang, Peng Zhang, Qinkai Zheng, Rui Lu, Shuaiqi Duan, Shudan Zhang, Shulin Cao, Shuxun Yang, Weng Lam Tam, Wenyi Zhao, Xiao Liu, Xiao Xia, Xiaohan Zhang, Xiaotao Gu, Xin Lv, Xinghan Liu, Xinyi Liu, Xinyue Yang, Xixuan Song, Xunkai Zhang, Yifan An, Yifan Xu, Yilin Niu, Yuantao Yang, Yueyan Li, Yushi Bai, Yuxiao Dong, Zehan Qi, Zhaoyu Wang, Zhen Yang, Zhengxiao Du, Zhenyu Hou, and Zihan Wang. 2024. Chatglm: A family of large language models from glm-130b to glm-4 all tools. *Preprint*, arXiv:2406.12793.

Global Times. 2023. China's aging population continues to rise as the country faces workforce challenges. *Global Times*. Accessed: 2024-09-17.

Alistair EW Johnson, Tom J Pollard, Lu Shen, Li-wei H Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. 2016. Mimic-iii, a freely accessible critical care database. *Scientific data*, 3(1):1–9.

LangChain. 2023. Langchain documentation. Accessed: 2023-09-17.

Binbin Li, Tianxin Meng, Xiaoming Shi, Jie Zhai, and Tong Ruan. 2023. Meddm: Llm-executable clinical guidance tree for clinical decision-making. *arXiv preprint arXiv:2312.02441*.

Rumeng Li, Xun Wang, and Hong Yu. 2024. Llamacare: An instruction fine-tuned large language model for clinical nlp. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 10632–10641.

Xiao Li. 2019. Medical postgraduate examination system analysis. *Journal of Medical Education*, 22:98–104.

Haotian Liu, Chunyuan Li, Yuheng Li, and Yong Jae Lee. 2023a. Improved baselines with visual instruction tuning.

Haotian Liu, Chunyuan Li, Yuheng Li, Bo Li, Yuanhan Zhang, Sheng Shen, and Yong Jae Lee. 2024. Llava-next: Improved reasoning, ocr, and world knowledge.

Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. 2023b. Visual instruction tuning.

Zabir Al Nazi and Wei Peng. 2024. Large language models in healthcare and medical domain: A review. In *Informatics*, volume 11, page 57. MDPI.

Du Peng. 2023. Negative population growth and population ageing in china. *China Population and Development Studies*, 7(2):95–103.

Nurs Educ Perspect. 2023. Dear editor: Large language models (llms), such as openai' s generative pretrained transformer series, are trained on vast amounts of text data and have demonstrated remarkable capabilities in understanding and generating human-like. *Journal of Emergency Nursing*, 49(5).

Yang Tan, Zhixing Zhang, Mingchen Li, Fei Pan, Hao Duan, Zijie Huang, Hua Deng, Zhuohang Yu, Chen Yang, Guoyang Shen, et al. 2024. Medchatzh: A tuning llm for traditional chinese medicine consultations. *Computers in Biology and Medicine*, 172:108290.

Arun James Thirunavukarasu, Darren Shu Jeng Ting, Kabilan Elangovan, Laura Gutierrez, Ting Fang Tan, and Daniel Shu Wei Ting. 2023. Large language models in medicine. *Nature medicine*, 29(8):1930–1940.

Ehsan Ullah, Anil Parwani, Mirza Mansoor Baig, and Rajendra Singh. 2024. Challenges and barriers of using large language models (llm) such as chatgpt for diagnostic medicine with a focus on digital pathology–a recent scoping review. *Diagnostic pathology*, 19(1):43.

Raja Vavekanand and Kira Sam. 2024. Llama 3.1: An in-depth analysis of the next-generation large language model.

Jun Wang. 2018. Evaluation metrics for nursing knowledge testing models. *Nursing Informatics Journal*, 18:45–53.

Weihan Wang, Qingsong Lv, Wenmeng Yu, Wenyi Hong, Ji Qi, Yan Wang, Junhui Ji, Zhuoyi Yang, Lei Zhao, Xixuan Song, Jiazheng Xu, Bin Xu, Juanzi Li, Yuxiao Dong, Ming Ding, and Jie Tang. 2023a. Cogvlm: Visual expert for pretrained language models. *Preprint*, arXiv:2311.03079.

Yan Wang, Heidi Donovan, Sabit Hassan, and Malihe Alikhani. 2023b. MedNgage: A dataset for understanding engagement in patient-nurse conversations. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 4613–4630, Toronto, Canada. Association for Computational Linguistics.

Brigitte Woo, Tom Huynh, Arthur Tang, Nhat Bui, Giang Nguyen, and Wilson Tam. 2024. Transforming nursing with large language models: from concept to practice. *European journal of cardiovascular nursing*, page zvad120.

Liping Xiong, Qiqiao Zeng, Wuhong Deng, Weixiang Luo, and Ronghui Liu. 2023. A novel approach to nursing clinical intelligent decision-making: Integration of large language models and local knowledge bases.

Yuexiang Zhai, Shengbang Tong, Xiao Li, Mu Cai, Qing Qu, Yong Jae Lee, and Yi Ma. 2023. Investigating the catastrophic forgetting in multimodal large language models. *arXiv preprint arXiv:2309.10313*.

Sifeng Zhang and Ze Zhang. 2023. Intelligent elderly care is the inevitable choice for china's elderly care services. *Xi'an Jiaotong University News*. Accessed: 2024-09-17.

Wei Zhang. 2020. Nursing examination model based on "three basics and three stricts". *Journal of Nursing Education*, 35:123–130.

Hongjian Zhou, Boyang Gu, Xinyu Zou, Yiru Li, Sam S Chen, Peilin Zhou, Junling Liu, Yining Hua, Chengfeng Mao, Xian Wu, et al. 2023. A survey of large language models in medicine: Progress, application, and challenge. *arXiv preprint arXiv:2311.05112*.

A Appendix

A.1 Details for the LangChain prompting.

LangChain (LangChain, 2023) is a powerful framework that enables developers to build applications powered by large language models (LLMs). It provides a suite of modular components, including Prompts, Indexes, Chains, Agents, and Memory, which developers can leverage to build a variety of intelligent applications such as personal assistants, question-answering systems, and chatbots. Furthermore, LangChain offers standardized interfaces, extensive integrations with third-party tools, and examples of common application use cases, allowing developers to more easily harness the capabilities of language models to construct their own tailored solutions.

This section provides detailed explanations and examples of how LangChain is utilized to implement various components of the dynamic nursing assistant system. Table 4 summarized the techniques combined in terms of modules and functions and below are the core components and corresponding LangChain implementations:

1. **Data Collection and Monitoring**: LangChain integrates with external tools to gather patient feedback and health data through natural language interfaces. It can process and format the input, converting it into structured data.

```
from langchain.prompts import
    PromptTemplate
from langchain.chains import LLMChain

template = """
收集以下健康数据:
- 心率
- 血压
- 患者主诉

输入: {user_input}
"""
prompt = PromptTemplate(template=
    template, input_variables=["
    user_input"])
chain = LLMChain(prompt=prompt)
result = chain.run("患者感觉头晕, 血压
    140/90, 心率90")
print(result)
```

2. **Triggering Nursing Diagnosis**: LangChain can automate nursing diagnosis by using rule-based engines or AI models, depending on patient health indicators.

```
from langchain.chains import
    SimpleSequentialChain

def check_for_issues(user_input):
    if "血压140/90" in user_input:
        return "触发高血压护理诊断"
    else:
        return "病情稳定"

def diagnostic_advice(issue):
    if "高血压" in issue:
        return "建议每日测量血压, 限制盐
            分摄入, 定期服用降压药"
    else:
        return "无特殊护理建议"

chain_1 = LLMChain(check_for_issues)
chain_2 = LLMChain(diagnostic_advice)

sequential_chain = SimpleSequentialChain
    (chains=[chain_1, chain_2])
result = sequential_chain.run("患者血压
    140/90, 心率90")
print(result)
```

3. **Personalized Care Plan Generation**: LangChain can generate personalized care plans by dynamically creating templates based on the patient's condition.

```
from langchain.prompts import
    PromptTemplate
```

```
from langchain.chains import LLMChain

template = """
患者状态: {user_input}
基于患者的状态，生成以下护理计划:
- 药物管理
- 饮食建议
- 康复计划

输入: {user_input}
"""
prompt = PromptTemplate(template=
    template, input_variables=["
    user_input"])
chain = LLMChain(prompt=prompt)
result = chain.run("高血压患者，血压140
    /90，心率90")
print(result)
```

4. **Continuous Monitoring and Feedback Adjustment**: LangChain allows for continuous patient feedback collection and care plan adjustments through persistent conversation chains.

```
from langchain.memory import
    ConversationBufferMemory
from langchain.chains import
    ConversationChain

memory = ConversationBufferMemory()
conversation = ConversationChain(memory=
    memory)

conversation.run("患者感觉心情好转，但仍
    有头晕")
conversation.run("继续测量血压并减少盐摄
    入")
conversation.run("血压已降至130/80，感觉
    良好")

print(memory.load_memory_variables({}))
```

5. **Dynamic Care Stage Transition**: LangChain can automatically assess patient status and trigger transitions between different stages of care based on health indicators.

```
def check_stage(patient_data):
    if "血压130/80" in patient_data:
        return "患者康复，进入后续健康管
            理阶段"
    else:
        return "继续当前护理"

chain_stage = LLMChain(check_stage)
result = chain_stage.run("患者血压130/80
    ，心率正常")
print(result)
```

6. **Health Education and Follow-Up Support**: LangChain can dynamically generate health education materials and reminders for patients in the recovery phase.

```
from langchain.prompts import
    PromptTemplate
```

```
from langchain.chains import LLMChain

template = """
患者恢复阶段: {user_input}
生成一份个性化的健康教育指南，帮助患者维
    持康复:
- 生活建议
- 饮食注意事项
- 每日健康监控任务

输入: {user_input}
"""
prompt = PromptTemplate(template=
    template, input_variables=["
    user_input"])
chain = LLMChain(prompt=prompt)
result = chain.run("患者进入康复阶段，血
    压正常")
print(result)
```

Module/Function	Description	Technology/Tools	Key Requirements
Data Collection and Monitoring	Collects patient health data (e.g., heart rate, blood pressure) and self-reported symptoms.	IoT devices, API integration (e.g., MQTT, HTTP/RESTful)	Ensures data is collected in real-time with high accuracy, reliable API integration.
Natural Language Data Processing	Processes patient-reported information and extracts key health data.	LangChain input-output chains, prompt templates	Accurate handling of input and non-standard language expressions.
Nursing Diagnosis Trigger	Triggers nursing diagnosis and generates recommendations based on collected data.	LangChain logic chains, AI diagnostic models	Utilizes rule-based engines or machine learning models in combination with external diagnostic APIs.
Personalized Care Plan Generation	Generates personalized care plans based on diagnostic results.	LangChain natural language generation (NLG)	Real-time updates and personalized care plans.
Continuous Monitoring and Feedback Adjustment	Continuously monitors patient status, collects feedback, and adjusts the care plan dynamically.	Stream processing (Kafka/Flink), LangChain memory chains	Efficient processing of sensor data, timely adjustments to the care plan.
Dynamic Care Stage Transition	Dynamically determines transitions between care stages based on patient recovery.	LangChain logic chains, state machines	Properly defined conditions for stage transitions using state machines or rule engines.
Health Education and Follow-Up Support	Provides post-care health education and periodic follow-up for patients.	LangChain NLG, messaging services	Dynamic generation of educational content and timely follow-up reminders.
Data Storage and Encryption	Encrypts and stores patient health data in a database.	AES-256 encryption, RSA encryption	Secure storage of encryption keys, ensuring data remains encrypted at all times.
Key Management and Access Control	Manages encryption keys securely through key management services.	AWS KMS, Google Cloud KMS	Implements key rotation and enforces strict access control policies.
Authentication and Key Access	Ensures access to sensitive data through authentication mechanisms.	OAuth 2.0, JWT	Prevents identity theft and ensures key security.

Table 4: Dynamic Nursing Assistant System Functional Modules

LoRA Parameters	
LoRA_alpha	24
LoRA_dropout	0.08
LoRA_rank	48
bias	None
Other Parameters	
num_train_epochs	4
per_device_train_batch_size	6
gradient_accumulation_steps	3
optimizer	paged_adamw
learning_rate	2.5e-4
tf32	True
max_grad_norm	0.4
warmup_ratio	0.02
max_length	4096
lr_scheduler_type	cosine
IPT Process Parameters	
num_ipt_epochs	3
pretrain_batch_size	12
learning_rate (IPT)	1.5e-4
max_grad_norm (IPT)	0.35
ipt_optimizer	adamw
warmup_steps	3000

Table 5: Parameters for Model Fine-tuning and IPT on 8x A100 80GB GPUs.