Proceedings of the Thirty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-26)
Demonstrations Track

RareDASH: A Dynamic Multi-Agent System for Holistic Rare Disease Care

Jialun Zhong[1] , **Jiayang Yu**[1] , **Yanzeng Li**[2] , **Meng Qin**[3] , **Lei Zou**[1] , **Yuqian Wang**[3] , **Ying Zhang**[3] , **Hanna Li**[3] , **Liying Yan**[3] and **Jie Qiao**[3]
[1]Wangxuan Institute of Computer Technology, Peking University
[2]Beijing Normal University
[3]Peking University Third Hospital
zhongjl@stu.pku.edu.cn, jyyu1010@foxmail.com, zoulei@pku.edu.cn

Abstract

Rare diseases are characterized by low prevalence and intricate pathogenesis, leading to highly heterogeneous clinical trajectories. The care of rare disease presents formidable challenges due to the requirement for highly specialized expertise and experiences. Existing methods are typically tailored for isolated rare disease scenarios (e.g., diagnostic tasks, medication recommendations), which lacks a comprehensive perspective of the entire care process. Inspired by recent studies of agent skills, we propose RareDASH, a multi-agent system (MAS) featuring dynamic workflow orchestration designed to provide a comprehensive solution for the full life-cycle of rare disease care. Our framework is inherently patient-centric, enhancing rare disease discovery capabilities through proactive inquiry and information elicitation directly from the patients. Furthermore, we implement diverse agent memory to optimize both the accuracy and efficiency of the multi-agent collaboration. Finally, an online auditing module is integrated into the system to monitor and mitigate the hallucinations, ensuring the reliability of clinical outputs. The work sheds light on the feasibility of leveraging MAS in holistic rare disease care.

1 Introduction

Rare diseases are typically characterized by low prevalence, complex pathogenesis, and life-threatening [Pistollato *et al.*, 2025]. Though individually infrequent, recent estimates indicate that approximately 7,000-10,000 rare diseases affect over 300 million people worldwide [Chaudhary and Kumar, 2025], posing a monumental challenge to global healthcare systems [Chung *et al.*, 2022]. Meanwhile, the full life-cycle of care, encompassing diagnosis, medication recommendation, and long-term psychosocial support is fragmented and resource-intensive. The scarcity of specialized expertise [Chou *et al.*, 2024] further exacerbates the disparity in care quality, frequently resulting in prolonged diagnostic odysseys and leaving a critical gap in the comprehensive management of these complex conditions.

Artificial Intelligence (AI) has emerged as a transformative force in health care [Zhi *et al.*, 2025; Tao *et al.*, 2026], clinical practice [Li *et al.*, 2024b; Zhong *et al.*, 2025a]. Substantial research [Wojtara *et al.*, 2023; Wang *et al.*, 2024; Zhao *et al.*, 2025a] has shown remarkable effectiveness in distinct phases of the rare disease care, catalyzing the digital transformation of this domain. Moreover, with the advancement of Large Language Model (LLM) technology, LLM-driven Multi-Agent Systems (MAS) [Li *et al.*, 2024a; Fan *et al.*, 2025] offer a promising paradigm to address the intricate demands and low-resource challenges in healthcare. Recent approaches [Chen *et al.*, 2024; Chen *et al.*, 2025; Zhao *et al.*, 2025b] have sought to integrate diverse medical tools and knowledge into MAS to bolster system robustness and clinical validity. However, while AI systems excel at specific tasks like diagnosis and medication recommendation, they often lack the holistic perspective and proactive coordination required for the full life-cycle of patient care.

To navigate the fragmented and resource-constrained landscape of rare disease care, we introduce **RareDASH**, a **D**ynamic Multi-**A**gent **S**ystem designed for **H**olistic Rare Disease Care. Our framework employs dynamic agent orchestration workflows, drawing inspiration from reusable agent skills [Xu and Yan, 2026] to enable modular collaboration, thereby addressing patient needs across diverse stages of the care continuum. To facilitate the early detection of potentially rare conditions, we incorporate a proactive inquiry mechanism that actively elicits crucial phenotypic information directly from patients. Recognizing that diagnoses are often time-consuming, we implement a reusable memory cache for agents, thereby improving communication efficiency and reducing redundant computation. Finally, to mitigate the risk of hallucination propagation within extended collaborative chains, we integrate a continuous auditing agent dedicated to the real-time monitoring of task execution and reasoning fidelity. Through a practical demonstration, we showcase the capability of RareDASH to manage the multifaceted care requirements of rare disease patients, offering a novel perspective for future research in this domain.

Our contributions include: (1) the first MAS RareDASH for holistic rare disease care, (2) four carefully designed modules for enhancing the clinical validity of the proposed system, (3) a user-friendly demonstration for rare disease patients that requires no specialized knowledge.

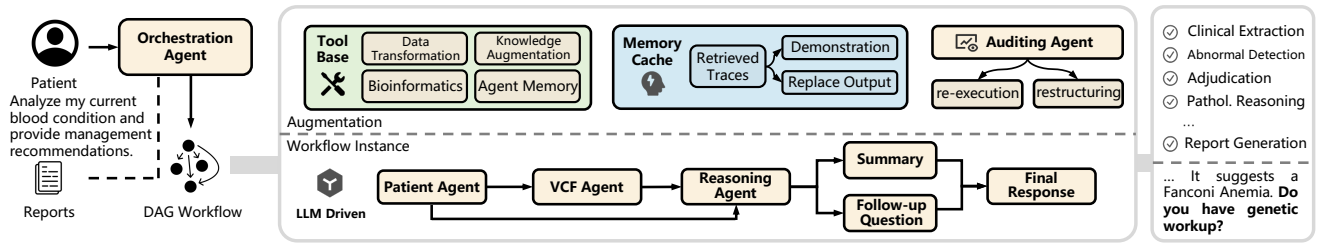

Figure 1: The framework of RareDASH. The patient's query is processed by the orchestration agent into a DAG workflow. The lower central panel illustrates a workflow instance, while the upper panel displays the augmentation modules. The final system response is synthesized by combining the workflow summary with a proactive follow-up question.

2 Methodology

In this section, we elaborate on the methodology of **RareDASH**, systematically detailing the four core modules that constitute the system, including dynamic orchestration workflow, proactive inquiry mechanism, memory cache, and continuous auditing.

Dynamic Orchestration Workflow. The full life-cycle of rare disease care encompasses a diverse array of clinical scenarios, where each scenario necessitates a distinct collaboration paradigm among specialized agents. Meanwhile, the low-resource characteristic of rare diseases compels the integration of heterogeneous external tools [Köhler *et al.*, 2009; Knox *et al.*, 2024] and domain-specific knowledge [1] to bridge information gaps, which significantly amplifies the complexity of system. Inspired by [Wei *et al.*, 2025], we leverage expert priors to construct predefined workflows structured as Directed Acyclic Graphs (DAGs) for common care scenarios to effectively manage this complexity. Formally, within this graph structure, each node represents an instantiated agent, while the directed edges denote the propagation of information between them. For diverse patient inquiries, these expert-curated workflows serve as high-quality references, empowering the orchestration agent to dynamically generate and tailor new workflows for novel or unseen inquiry types based on the specific context. This topological constraint is critical for preventing the MAS from descending into infinite communication loops or stagnating at sub-optimal local solutions.

Proactive Inquiry Mechanism. Passive reliance on patient descriptions often leads to diagnostic problems, particularly given the characteristic of imperfect information on the user side [Zhang *et al.*, 2025a; Li *et al.*, 2026] and phenotypic heterogeneity in rare diseases. To counteract this, we introduce a proactive inquiry module [Mulla and Gharpure, 2023] designed to transition the system from a passive responder to an active investigator. This mechanism empowers agents to dynamically formulate and pose targeted follow-up questions when the available information is insufficient for a confident conclusion. By iteratively eliciting latent patient-specific information and clarifying vague conditions directly from the patients, the system significantly enhances the granularity of the clinical context, thereby facilitating earlier and more accurate rare disease discovery.

[1] https://pubmed.ncbi.nlm.nih.gov/

Memory Cache. The computational cost and temporal latency associated with complex multi-agent collaboration chains can impede real-time clinical interaction, especially those steps involving extensive tool calling. To address this efficiency bottleneck, we implement a collaborative memory cache for each agent that persists original input, reasoning trajectories and tool execution results. Formally, this module persists every successful agent execution into a structured memory cache for each agent. Upon encountering recurring sub-tasks or analogous clinical queries, the system retrieves the most semantically similar trajectories from the memory. When the retrieved input of one trajectory exhibit an exact match with the current request, the system directly serves the validated output to the user rather than re-computes them from scratch. Conversely, when an exact match is absent, the retrieved trajectories are utilized as in-context demonstrations to guide reasoning. Crucially, to uphold data privacy, memory caches are strictly isolated across distinct patient identities. Ultimately, this mechanism anchors agent reasoning to verified traces and significantly mitigates latency, especially for continuous patient care.

Continuous Auditing. The collaboration chains in MAS are susceptible to the propagation of hallucinations [Zhang *et al.*, 2025b], where a minor error in an early reasoning stage can cascade into severe clinical problems. We establish a Continuous Auditing Agent that operates in parallel with the primary workflow execution to enforce clinical rigor. Adopting the LLM-as-a-Judge [Gu *et al.*, 2024; Zhong *et al.*, 2025b] strategy, this module checks intermediate reasoning steps and tool outputs against established hallucination types, medical constraints and instruction consistency. If a potential deviation is detected, the audit agent immediately intervenes to trigger a correction loop and generates revision suggestions, effectively mitigating the risk of hallucination and ensuring the safety and fidelity of the final responses.

3 System Design & Architecture

Following the conceptualization of key modules for holistic care in Section 2, we implemented RareDASH by synthesizing these modules into a unified multi-agent environment. The schematic architecture is shown in Figure 1, with a corresponding practical demonstration displayed in Figure 2.

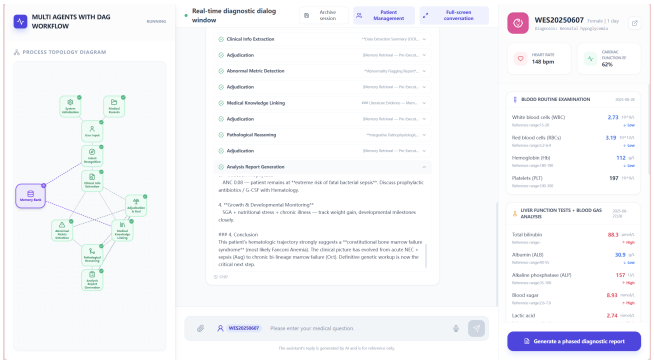

Figure 2: The demonstration for RareDASH. Left: The DAG of the current workflow. Center: The dialogue interface, where each agent response is visible. Right: The Information of the current patient.

Orchestration Agent. Serving as the central unit, the orchestration agent is responsible for parsing the initial patient query and determining the optimal execution pathway. Leveraging the expert-curated DAGs described in Section 2, it dynamically instantiates the necessary sub-agents and schedules their interactions. By mapping unstructured patient requests to structured clinical workflows (e.g., symptom checking vs. medication consultation), this agent ensures that the system's operation aligns with established medical protocols rather than relying on unstructured, free-form generation.

Patient-Specific Agent. This agent is dedicated to the parsing and management of patient-centric data. It is equipped with multi-modal parsing capabilities to process diverse uploaded medical documents, such as clinical reports, medical imaging, and electronic health records. By leveraging information extraction tools like Optical Character Recognition (OCR), the agent parses these heterogeneous files to text-format report for downstream agents. Furthermore, it is responsible for continuously maintaining the dynamic context of the patient's status.

Variant Call Format Agent. Variant Call Format (VCF) is a standardized format used in bioinformatics to store gene sequence variations. This agent is purposed to handle the complexities of **genomic** data. It is responsible for executing VCF-related directives, such as pathogenicity analysis and genotype-based diagnosis, serving as a complementary method for phenotype-based analysis.

Reasoning Agent. This agent is responsible for carrying out the core clinical operations. Its scope encompasses critical tasks such as differential diagnosis, medication recommendation, and post-intervention recommendation. Following the guidance of the orchestration agent, it executes the assigned workflow steps in a sequential manner. Crucially, the agent generates intermediate reasoning steps in the verifiable text format to ensure system interpretability and facilitate human validation.

Auditing Agent. Instantiating the Continuous Auditing module formalized in Section 2, the agent is dedicated to the real-time detection of failures or hallucinations within the execution workflow. The detected errors are intercepted before propagating to the final output. Depending on the severity and categories of the error, the agent will manipulate the underlying Directed Acyclic Graph (DAG), ranging from the re-execution of specific nodes to the dynamic restructuring of the graph. It effectively steers the reasoning process back toward a valid clinical trajectory.

Module	Model / Technology
LLM Backbone	Qwen3-Next-80B-Instruct (API)
Phenotype Knowledge	HPO Database (SQLite) + FAISS
Phenotype Embedding	SapBERT (biomedical-trained)
Literature RAG	PubMed + MedCPT Dual-Encoder
Web Retrieval	Google Search API
Framework	Hierarchical Multi-Agent System

Table 1: Core System Components and Implementation.

Tool Integration. To tackle the limitations of static parametric rare disease knowledge within LLMs, we integrate external tools divided into four categories:

- **Heterogeneous Data Transformation.** Reports scanning, PDF parsing, structured tabular data extraction, free-text HPO terms extraction, etc.
- **Knowledge Augmentation.** Public knowledge sources (e.g., PubMed, HPO, OMIM, Bing API), in-house databases (e.g., Patient cases, knowledge graphs).
- **Biocomputational Tools.** Pathogenicity assessment, transcript matching, nonsense-mediated decay prediction, amino acid substitution effect, etc.
- **Agent Memory.** Trajectories retrieval and update.

4 Implementation

The backend of RareDASH is implemented in Python with an asynchronous event-driven architecture, serving agent responses to the frontend via a WebSocket interface that supports real-time streaming output. The core components and their underlying technologies are listed in Table 1.

5 Conclusions

In this paper, we presented RareDASH, a dynamic multi-agent system designed to address the fragmented and resource-constrained landscape of holistic rare disease care. By integrating four complementary modules, RareDASH transcends the limitations of task-specific AI systems and provides a principled framework for managing the full lifecycle of rare disease care. Through a practical multi-scenario demonstration spanning medical consultation, phenotype extraction, genomic variant analysis, and multi-modal clinical data integration, we have showcased the breadth and adaptability of RareDASH in realistic rare disease care settings. Looking forward, we aim to conduct a rigorous evaluation of RareDASH against established benchmarks and to explore its deployment in real-world hospital environments, with the goal of further uncovering new challenges unique to the holistic care of rare disease patients.

Ethical Statement

This study was approved by the Institutional Review Board of Peking University Third Hospital (Approval No. IRB00006761-M20251114) and conducted in accordance with the Declaration of Helsinki and relevant ethical guidelines.

Acknowledgments

This work was supported by New Generation Artificial Intelligence-National Science and Technology Major Project under grant 2025ZD0123304, NSFC under grant 62532001, and Beijing Research Ward Excellence Program under grant BRWEP2024W094090108. Jialun Zhong and Jiayang Yu contributed equally to this work. The corresponding author of this work is Lei Zou.

References

[Chaudhary and Kumar, 2025] Aryan Chaudhary and Vidyapati Kumar. Rare diseases: a comprehensive literature review and future directions. *Journal of Rare Diseases*, 4(1):33, 2025.

[Chen *et al.*, 2024] Xuanzhong Chen, Ye Jin, Xiaohao Mao, Lun Wang, Shuyang Zhang, and Ting Chen. Rareagents: Autonomous multi-disciplinary team for rare disease diagnosis and treatment. *arXiv preprint arXiv:2412.12475*, 2024.

[Chen *et al.*, 2025] Xi Chen, Huahui Yi, Mingke You, WeiZhi Liu, Li Wang, Hairui Li, Xue Zhang, Yingman Guo, Lei Fan, Gang Chen, et al. Enhancing diagnostic capability with multi-agents conversational large language models. *NPJ digital medicine*, 8(1):159, 2025.

[Chou *et al.*, 2024] Cody Chou, Sydney O Wiredu, Liesel Von Imhof, Anran Tan, Sasha Agarwal, Melis Lydston, and Vanessa L Merker. Health services interventions to improve the quality of care in rare disease: a scoping review. *Therapeutic Advances in Rare Disease*, 5:26330040241262810, 2024.

[Chung *et al.*, 2022] Claudia Ching Yan Chung, Hong Kong Genome Project, Annie Tsz Wai Chu, and Brian Hon Yin Chung. Rare disease emerging as a global public health priority. *Frontiers in public health*, 10:1028545, 2022.

[Fan *et al.*, 2025] Zhihao Fan, Lai Wei, Jialong Tang, Wei Chen, Wang Siyuan, Zhongyu Wei, and Fei Huang. Ai hospital: Benchmarking large language models in a multi-agent medical interaction simulator. In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 10183–10213, 2025.

[Gu *et al.*, 2024] Jiawei Gu, Xuhui Jiang, Zhichao Shi, Hexiang Tan, Xuehao Zhai, Chengjin Xu, Wei Li, Yinghan Shen, Shengjie Ma, Honghao Liu, et al. A survey on llm-as-a-judge. *The Innovation*, 2024.

[Knox *et al.*, 2024] Craig Knox, Mike Wilson, Christen M Klinger, Mark Franklin, Eponine Oler, Alex Wilson, Allison Pon, Jordan Cox, Na Eun Chin, Seth A Strawbridge, et al. Drugbank 6.0: the drugbank knowledgebase for 2024. *Nucleic acids research*, 52(D1):D1265–D1275, 2024.

[Köhler *et al.*, 2009] Sebastian Köhler, Marcel H Schulz, Peter Krawitz, Sebastian Bauer, Sandra Dölken, Claus E Ott, Christine Mundlos, Denise Horn, Stefan Mundlos, and Peter N Robinson. Clinical diagnostics in human genetics with semantic similarity searches in ontologies. *The American Journal of Human Genetics*, 85(4):457–464, 2009.

[Li *et al.*, 2024a] Yanzeng Li, Cheng Zeng, Jinchao Zhang, Jie Zhou, and Lei Zou. Meddit: A knowledge-controlled diffusion transformer framework for dynamic medical image generation in virtual simulated patient. *arXiv preprint arXiv:2408.12236*, 2024.

[Li *et al.*, 2024b] Yanzeng Li, Cheng Zeng, Jialun Zhong, Ruoyu Zhang, Minhao Zhang, and Lei Zou. Leveraging large language model as simulated patients for clinical education. *arXiv preprint arXiv:2404.13066*, 2024.

[Li *et al.*, 2026] Yanzeng Li, Xiaoning Cao, Jialun Zhong, Jianpeng Hu, Jiangshan Tan, Ningning Liu, Feng Xiang, and Shasha Han. Dysrec: Dynamic context-aware psychometric scale recommendation via multi-agent collaboration. *arXiv preprint arXiv:2605.00574*, 2026.

[Mulla and Gharpure, 2023] Nikahat Mulla and Prachi Gharpure. Automatic question generation: a review of methodologies, datasets, evaluation metrics, and applications. *Progress in Artificial Intelligence*, 12(1):1–32, 2023.

[Pistollato *et al.*, 2025] Francesca Pistollato, Fabia Furtmann, Lindsay J Marshall, Surat Parvatam, Jan Turner, Flora Tshinanu Musuamba, Giulia Russo, and Francesco Pappalardo. Advancing the frontier of rare disease modeling: a critical appraisal of in silico technologies. *npj Digital Medicine*, 8(1):676, 2025.

[Tao *et al.*, 2026] Xinge Tao, Shuya Zhou, Kai Ding, Sairan Li, Yanzeng Li, Boyou Wu, Qirui Huang, Wangyue Chen, Muzi Shen, En Meng, et al. An llm chatbot to facilitate primary-to-specialist care transitions: a randomized controlled trial. *Nature Medicine*, pages 1–9, 2026.

[Wang *et al.*, 2024] Guanchu Wang, Junhao Ran, Ruixiang Tang, Chia-Yuan Chang, Yu-Neng Chuang, Zirui Liu, Vladimir Braverman, Zhandong Liu, and Xia Hu. Assessing and enhancing large language models in rare disease question-answering. *arXiv preprint arXiv:2408.08422*, 2024.

[Wei *et al.*, 2025] Xiaolong Wei, Yuehu Dong, Xingliang Wang, Xingyu Zhang, Zhejun Zhao, Dongdong Shen, Long Xia, and Dawei Yin. Beyond react: A planner-centric framework for complex tool-augmented llm reasoning. *arXiv preprint arXiv:2511.10037*, 2025.

[Wojtara *et al.*, 2023] Magda Wojtara, Emaan Rana, Taibia Rahman, Palak Khanna, and Heshwin Singh. Artificial intelligence in rare disease diagnosis and treatment. *Clinical and Translational science*, 16(11):2106–2111, 2023.

[Xu and Yan, 2026] Renjun Xu and Yang Yan. Agent skills for large language models: Architecture, acqui-

sition, security, and the path forward. *arXiv preprint arXiv:2602.12430*, 2026.

[Zhang *et al.*, 2025a] Bolin Zhang, Shengwei Wang, Yangqin Jiang, Dianbo Sui, Zhiying Tu, and Dianhui Chu. Ask and retrieve knowledge: Towards proactive asking with imperfect information in medical multi-turn dialogues. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1055–1065, 2025.

[Zhang *et al.*, 2025b] Shaokun Zhang, Ming Yin, Jieyu Zhang, Jiale Liu, Zhiguang Han, Jingyang Zhang, Beibin Li, Chi Wang, Huazheng Wang, Yiran Chen, et al. Which agent causes task failures and when? on automated failure attribution of llm multi-agent systems. *arXiv preprint arXiv:2505.00212*, 2025.

[Zhao *et al.*, 2025a] Chuang Zhao, Hui Tang, Jiheng Zhang, and Xiaomeng Li. Unveiling discrete clues: Superior healthcare predictions for rare diseases. In *Proceedings of the ACM on Web Conference 2025*, pages 1747–1758, 2025.

[Zhao *et al.*, 2025b] Weike Zhao, Chaoyi Wu, Yanjie Fan, Xiaoman Zhang, Pengcheng Qiu, Yuze Sun, Xiao Zhou, Yanfeng Wang, Xin Sun, Ya Zhang, et al. An agentic system for rare disease diagnosis with traceable reasoning. *arXiv preprint arXiv:2506.20430*, 2025.

[Zhi *et al.*, 2025] Xiaoquan Zhi, Hongke Zhao, Likang Wu, Chuang Zhao, and Hengshu Zhu. Reinventing clinical dialogue: Agentic paradigms for llm enabled healthcare communication. *arXiv preprint arXiv:2512.01453*, 2025.

[Zhong *et al.*, 2025a] Jialun Zhong, Yanzeng Li, Sen Hu, Yang Zhang, Teng Xu, and Lei Zou. Gap: Graph-assisted prompts for dialogue-based medication recommendation. *arXiv preprint arXiv:2505.12888*, 2025.

[Zhong *et al.*, 2025b] Jialun Zhong, Wei Shen, Yanzeng Li, Songyang Gao, Hua Lu, Yicheng Chen, Yang Zhang, Wei Zhou, Jinjie Gu, and Lei Zou. A comprehensive survey of reward models: Taxonomy, applications, challenges, and future. *arXiv preprint arXiv:2504.12328*, 2025.