

Original Paper

Immersive, Interactive, Intelligent Patient Educational System for Venous Thromboembolism (ChatVTE): Development and Validation Study

Bin bin Liu^{1*}, MM; Zhi geng Jin^{2*}, MD; Zhe qi Zhang¹, MM; Hong Wang¹, MS; Hao Wang³, MD; Hui Zhang², MD; Chang zhen Li⁴, MS; Fei Qi⁴, MD, PhD; Yu tao Guo², MD, PhD

¹Chinese People's Liberation Army Medical School, Chinese People's Liberation Army General Hospital, Beijing, China

²Department of Pulmonary Vascular and Thrombotic Disease, The Sixth Medical Center of Chinese People's Liberation Army General Hospital, Beijing, China

³Department of Cardiology, The Second Medical Center of Chinese People's Liberation Army General Hospital, Beijing, China

⁴Technical Department, DrBreath Medical Technology Co., Ltd., Beijing, China

*these authors contributed equally

Corresponding Author:

Yu tao Guo, MD, PhD

Department of Pulmonary Vascular and Thrombotic Disease

The Sixth Medical Center of Chinese People's Liberation Army General Hospital

6 Fucheng Road, Haidian District

Beijing 100048

China

Phone: 86 0-106-695-7703

Email: dor_guoyt@hotmail.com

Abstract

Background: Effective patient education is crucial in preventing venous thromboembolism (VTE), improving patient outcomes, and reducing health care costs. However, traditional educational methods often lack engagement and fail to address individual patient needs comprehensively.

Objective: This study aimed to develop and preliminarily validate an immersive, large language model-based patient education system for VTE designed to promote patient engagement and care adherence by delivering highly relevant, actionable, and patient-centered information.

Methods: We developed ChatVTE, an interactive, intelligent patient education platform, by integrating a retrieval-augmented large language model (Qwen1.5-7B) with text-to-speech and lip-synch technologies. The system's performance was initially assessed through a comparative evaluation against ChatGPT. This involved using a standardized set of VTE-related questions, administered from December 10 to 31, 2024, with responses rigorously evaluated by 4 VTE domain experts using a 5-point Likert scale for accuracy, completeness, consistency, and safety. Subsequently, we consecutively enrolled a prospective cohort of 25 adult inpatients with VTE from the Departments of Pulmonary Vascular and Thrombotic Diseases and General Surgery at the Sixth Medical Center of the Chinese People's Liberation Army General Hospital between March 1 and May 31, 2025. These participants engaged with the ChatVTE system throughout their inpatient stay and completed postintervention assessments upon discharge.

Results: Expert evaluation demonstrated that ChatVTE significantly outperformed ChatGPT in accuracy, completeness, consistency (all $P < .001$, $r > 0.5$), and safety ($P = .01$, $r = 0.327$). Among the 25 enrolled patients (age: mean 55.4, SD 13.2 years), ChatVTE achieved high average scores (mean score $> 4.0/5.0$) in 8 of the 9 experience dimensions evaluated but received a notably lower score in the emotional support domain (1.92/5.0).

Conclusions: This study validates the feasibility of ChatVTE in the management of patients with VTE, demonstrating its potential to enhance the quality of patient-health care provider interaction and the efficacy of knowledge dissemination. These preliminary findings suggest that ChatVTE could be a valuable tool for improving patient education and facilitating shared clinical decision-making.

Keywords: venous thromboembolism; large language model; retrieval-augmented generation; digital health; patient education; psychosocial; patient-centered care

Introduction

Venous thromboembolism (VTE), encompassing deep venous thrombosis and pulmonary embolism, is a predominant cause of avoidable mortality and morbidity on a global scale, ranking among the top 3 cardiovascular emergencies in high-income nations [1]. Despite progress in prophylactic strategies, a significant gap remains in patient education, which is a fundamental aspect of sustained VTE management [2]. Current evidence-based interventions largely concentrate on the optimization of anticoagulant therapy. However, recent research emphasizes the critical importance of education in improving adherence to prophylaxis, recognizing symptoms, and reducing the risk of recurrent hospitalizations [3-5]. Patients exhibit a significant interest in acquiring knowledge about VTE through various educational modalities [6]. Nonetheless, conventional educational approaches, which typically involve clinician-led verbal instruction and static written materials, are limited by their one-way nature and lack of continuity.

Mobile health technologies, such as smartphone apps, have emerged as a promising means for delivering dynamic and personalized educational content [7]. Nevertheless, many current mobile health platforms use rule-based content delivery systems, which fail to address patients' evolving needs. Large language models (LLMs), such as GPT-4, can generate contextual responses, offering a promising avenue for personalized patient education [8-10]. However, this potential is constrained by the tendency of general-domain models to produce inaccurate or outdated medical information. Retrieval-augmented generation (RAG) architectures address these limitations by grounding responses in real-time access to authoritative medical databases and clinical guidelines [11], a method validated in prior clinical applications [12,13].

Building upon recent technological advancements, we previously developed and validated a mobile venous thromboembolism app (mVTEA), a mobile app for VTE management. Its core functionalities, including risk stratification, medication adherence monitoring, and teleconsultation services, have been shown to enhance adherence to prophylaxis [14]. However, user feedback identified significant limitations in educational interactivity and personalized engagement. To address these challenges, we designed

ChatVTE, an immersive intelligent system integrating LLMs with mVTEA's existing framework. This system uses RAG architecture to dynamically synthesize evidence-based responses from curated clinical guidelines and patient education repositories, enabling real-time, context-aware interactions addressing complex clinical inquiries. This study aimed to develop and preliminarily validate an immersive LLM-driven patient education system for VTE.

Methods

Study Design and Setting

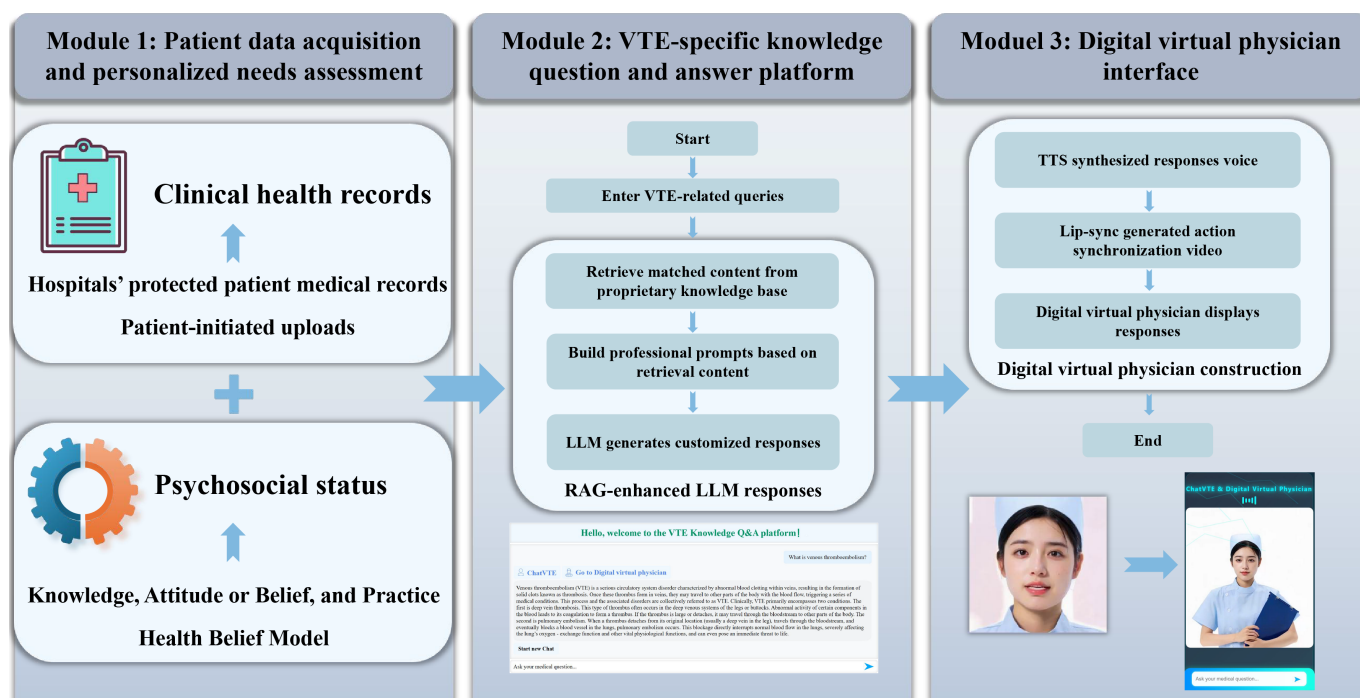
This study was designed as a pilot study to evaluate the feasibility and user experience of ChatVTE in a real-world clinical setting. It was conducted using a multiphase approach: an initial technical validation phase involving expert evaluation, followed by a prospective single-arm cohort study assessing patient experience. Both phases were conducted in a tertiary medical center in Beijing, China, specifically involving patients from the Departments of Pulmonary Vascular and Thrombotic Diseases and General Surgery at the Sixth Medical Center of the Chinese People's Liberation Army General Hospital. This study adhered to the iCHECK-DH (Guidelines for the Reporting on Digital Health Implementations; [Checklist 1](#)) [15].

ChatVTE Platform Engineering

Overview

ChatVTE is an interactive, intelligent patient education platform for VTE, developed by integrating a retrieval-augmented LLM (Qwen1.5-7B), text-to-speech (TTS) synthesis, and lip-synch technologies. This platform accepts VTE-related queries, retrieves validated content from a dedicated knowledge repository, processes the information via RAG technology, and then delivers a tailored response. The development process systematically constructed 3 interconnected core modules ([Figure 1](#)). Through this integration, the platform provides patients with personalized VTE education content, self-management strategies, and guidance on health-promoting behaviors. The core purpose is to deliver highly relevant, actionable, and patient-centered information, thereby promoting patient engagement and adherence to care.

Figure 1. ChatVTE architecture: an immersive retrieval-augmented generation (RAG) large language model (LLM) patient education platform for venous thromboembolism (VTE) integrating data capture, personalized needs assessment, a knowledge question and answer engine, and a virtual clinician. TTS: text-to-speech.



Module 1: Patient Data Acquisition and Personalized Needs Assessment Module

This module supports comprehensive patient profiling by securely integrating data from clinical records and patient-reported questionnaires. Clinical data were digitized from patient-authorized clinical documents using an optical character recognition (OCR)-based extraction process and organized into structured clinical entities to support personalized patient education and needs assessment [16]. Additionally, the module deploys periodic, structured questionnaires via the mVTEA platform to evaluate patients' knowledge, attitude, and practice (KAP) profiles as well as health belief model (HBM) domains pertinent to VTE. Questionnaire responses are securely consolidated and analyzed to quantify health-literacy levels, delineate behavioral patterns, and detect knowledge gaps or adherence barriers. Integrating these psychosocial metrics with the clinically derived data enables precise characterization of each patient's needs and informs the generation of tailored educational material and intervention strategies within ChatVTE.

Module 2: VTE-Specific Knowledge Question and Answer Platform

This module serves as the central intelligence for delivering evidence-based VTE information. A comprehensive knowledge repository was meticulously curated from authoritative open-access internet resources, including professional society guidelines (eg, from the European Society of Cardiology and the American Heart Association), peer-reviewed literature, and established biomedical databases such as PubMed. The aggregated material encompassed the full spectrum of VTE topics, including etiology, risk factors, clinical presentation, diagnostic

techniques, pharmacologic and nonpharmacologic management, prevention, and long-term care strategies. All retrieved data were subjected to a stringent preprocessing workflow that eliminated obsolete entries, harmonized medical terminology, and verified factual accuracy. The refined corpus was partitioned into discrete knowledge units and converted into vector representations. These embeddings enable a RAG engine to rapidly retrieve contextually relevant content for the Qwen1.5-7B LLM. The model subsequently integrates the extracted evidence to produce precise, context-sensitive, and patient-oriented textual responses. To prevent misinterpretation of system outputs as clinical orders, ChatVTE was explicitly designed as a patient education system rather than a clinical decision or order-entry tool. The interface includes clear disclaimers, and the system does not generate executable clinical instructions, thereby reducing the risk of misinterpretation as medical orders.

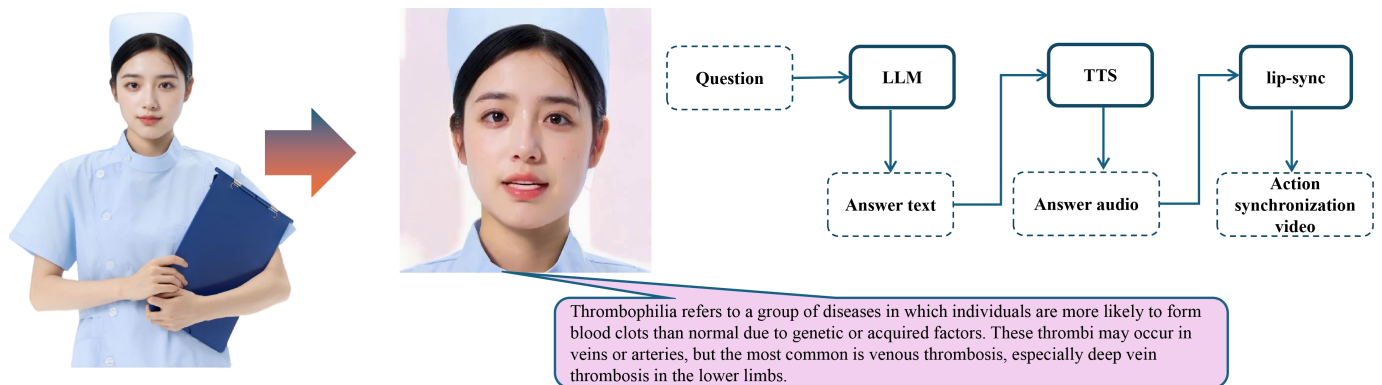
Module 3: Digital Virtual Physician Interface

This module facilitates an immersive and naturalistic communication experience. It transforms the textual output from the VTE-specific knowledge question and answer platform (module 2) into professional medical-style streaming audio via a specialized TTS module (Figure 2). The TTS module uses a pretrained medical voice corpus to ensure clear articulation of complex terminology, appropriate intonation for empathy and authority, and precise pronunciation. Concurrently, lip-synch technology generates precisely synchronized lip movements and microexpressions for the digital virtual clinician. This system maps speech segments to visual animations in real time, achieving a rendering rate of 24 to 30 frames per second with asynchronous latency of under 50 milliseconds. This process overlays the animations onto the virtual clinician's facial model,

significantly improving the visual realism and engagement of the interactive presentation. For example, we posed a query “What is venous thromboembolism?” to ChatVTE.

The resulting streaming video is provided in [Multimedia Appendix 1](#).

Figure 2. Digital virtual physician: patient education. LLM: large language model; TTS: text-to-speech.



The digital virtual physician, with the latest speech synthesis, image-based video, and lip-sync technologies, improved the quality of physician-patient interaction and enhanced the effectiveness of knowledge dissemination.

Evaluation Protocol

A dual-phase evaluation protocol was implemented to assess the system’s technical output quality and its practical utility for patients.

Expert Comparative Validation

We rigorously benchmarked the performance of ChatVTE against ChatGPT (GPT-4) between December 10 and 31, 2024. GPT-4 was selected as the high-performance general-purpose LLM baseline to assess the added value conferred by ChatVTE’s domain-specific optimization. We accessed GPT-4 via its official web interface ChatGPT, with web search disabled to ensure responses derived solely from its internal knowledge. Default safety and content moderation settings were applied without modification. This configuration was designed to simulate a typical scenario in which an end user directly interacts with a general-purpose LLM to seek medical information, thereby establishing a clear baseline for evaluating the added value of ChatVTE’s specialized design.

Both models responded to a standardized set of 30 VTE-related questions, which were carefully selected from the mVTEA education section, and comprehensively spanned

key management domains (Table S1 in [Multimedia Appendix 2](#)). These included pathophysiology; risk factors; diagnosis and clinical classification; acute and long-term treatment strategies; care for special populations (cancer, pregnancy, etc), and essential patient education for self-management and follow-up. Each of the 30 questions was posed to both ChatVTE and ChatGPT on 3 separate occasions: an initial round and 2 repeat rounds at 72-hour intervals, using semantically equivalent but differently worded formulations. Thus, a total of 180 individual responses were generated (2 models×30 questions×3 generations). The initial round of responses was used for evaluating accuracy, completeness, and safety; all 3 rounds were considered for the consistency assessment. Responses were independently assessed by 4 VTE domain experts (each possessing ≥7 years of clinical experience), who were blinded to the model source ([Text-box 1](#)). Responses were rated using 5-point Likert scales for accuracy, completeness, and consistency (1: very inaccurate, incomplete, or inconsistent; 5: very accurate, complete, or consistent). Safety, defined as the potential risk posed by erroneous or misleading information, was assessed on a separate 5-point severity scale (0: no risk; 4: extreme risk). Detailed anchor definitions are provided in Table S2 in [Multimedia Appendix 2](#).

Textbox 1. Summary of assessment area, measurement tool, and primary outcome for ChatVTE evaluation.

Expert-comparative performance vs ChatGPT (these items were scored in a blinded review on a 1-5 Likert scale, with a higher score indicating a better evaluation)

- accuracy
- completeness
- consistency
- safety (this item was rated on a separate 0-4 Likert scale, with a lower score indicating greater safety)

Patient-overall experience (these items were scored on a 1-5 Likert scale, with a higher score indicating a better evaluation)

- acceptability
- convenience

- timeliness
- fluency
- comprehensibility
- accuracy
- empathy
- satisfaction
- recommendability

Patient Experience Assessment

From March 1 to May 31, 2025, we consecutively enrolled 25 adult inpatients with a confirmed VTE diagnosis into a single-arm, single-center pilot study. Participants were provided with access to and guidance on using the ChatVTE system throughout their hospitalization. Upon discharge, each individual completed a structured 9-item questionnaire, expanded from earlier mVTEA evaluations, to preliminarily assess their overall experience (Table S3 in [Multimedia Appendix 2](#)). The questionnaire measured 9 dimensions: acceptability, convenience, timeliness, fluency, comprehensibility, accuracy, empathy, satisfaction, and recommendability, using a 5-point Likert scale (higher scores indicate more favorable evaluations; [Textbox 1](#)).

To ensure the instrument's suitability for this study's context, we used an iterative development and qualitative validation process. Prior to formal data collection, a qualitative pilot test was conducted with a small group of hospitalized patients with VTE ($n=5$). The primary objective was to evaluate the instrument's face validity and content validity, specifically focusing on the clarity of terminology, ease of comprehension, and respondent burden. During this pilot phase, we conducted brief cognitive debriefings to identify any ambiguous phrasing. On the basis of patient feedback, we linguistically refined several items to ensure that they were culturally and clinically appropriate for the target population. While formal psychometric analysis (eg, Cronbach α) was not performed at this stage due to the small sample size of the pilot group, this iterative process ensured that the questionnaire was content-valid and optimized for the clinical study.

Ethical Considerations

This study involved human participants and was approved by the ethics committees of the Sixth Medical Center of the Chinese People's Liberation Army General Hospital (HZKY-PJ-2022-21). Written informed consent was obtained from all participants before their inclusion in the study. Participants were informed of the purpose of the study, the nature of their interaction with the ChatVTE, and their right to withdraw at any time without consequences. No compensation was provided for participation. Clinical documents and inpatient records were acquired through secure, hospital-approved channels, including patient-authorized document upload and OCR-based extraction [16]. OCR processing was performed within a controlled hospital environment. Direct personal identifiers were separated from clinical content immediately after extraction, and only deidentified limited datasets were used for downstream processing. Identifiable information was retained locally and was not transmitted to or processed by

the language model. Access to patient data was restricted to authorized users through role-based authentication, and data access and processing events were logged to support auditability. All data were stored on secure in-country servers in accordance with institutional and national data protection regulations.

To protect participant privacy and data confidentiality, ChatVTE did not support real-time, bidirectional integration with the hospital information system during this pilot phase.

Statistical Analysis

Previous methodological work has demonstrated that pilot studies are appropriately sized to detect feasibility or implementation-related problems rather than to achieve statistical significance. Using a problem-detection framework, a sample size in the range of approximately 20 to 30 participants is sufficient to identify, with 95% confidence, feasibility or workflow-related issues that occur with a probability of at least 10% to 15% among study participants [17]. Given that the primary objectives of this study were to evaluate system feasibility and user experience in a relatively homogeneous patient population, a target sample size of 25 participants was considered adequate to meet these exploratory aims and to inform the design of subsequent larger-scale studies.

Quantitative data are summarized descriptively. Continuous variables are presented as mean (SD) or median (IQR). Expert evaluation scores for each model were calculated as the average ratings per question. Given that the same 30 questions were evaluated for both ChatVTE and ChatGPT, between-model comparisons were performed using the paired Wilcoxon signed-rank test. Effect sizes were quantified using the r statistic, calculated as $r = Z/\sqrt{N}$, where Z is the test statistic from the Wilcoxon signed-rank test. On the basis of the Cohen convention, an r value of approximately 0.1 indicates a small effect, approximately 0.3 indicates a medium effect, and $r \geq 0.5$ indicates a large effect [18,19]. The sign of Z (and thus r) indicates the direction of the difference: a positive value indicates that ChatVTE received higher scores than ChatGPT, and a negative value indicates the opposite. The absolute value of r reflects the magnitude of the effect. Patient evaluation scores were determined by the average score for each question. A 2-sided $P < .05$ was considered statistically significant. All statistical analyses were performed using SPSS Statistics (version 26.0; IBM Corp) and R (version 4.4.1; R Foundation for Statistical Computing).

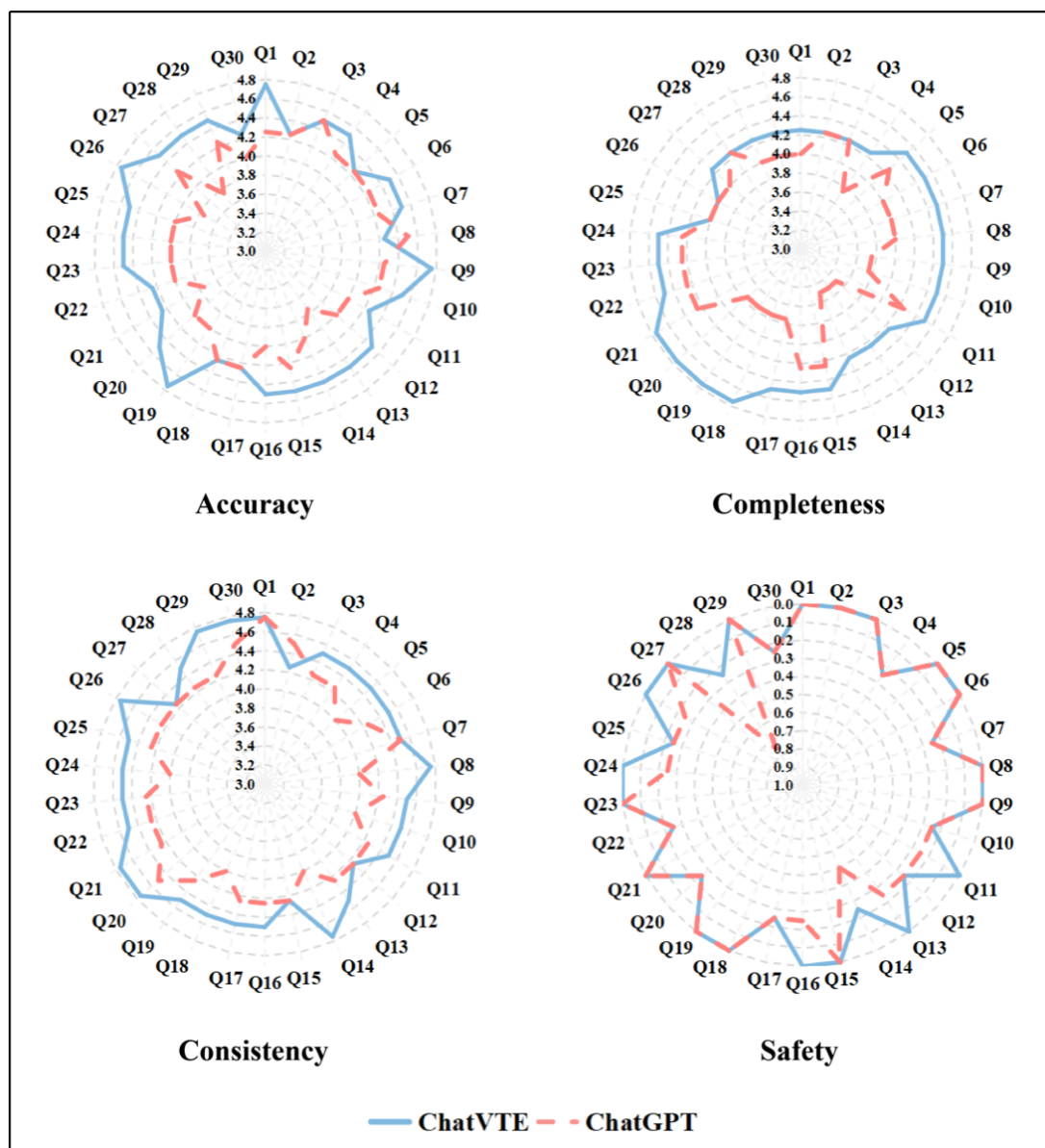
Results

Expert Comparative Validation Findings

Four VTE experts independently appraised all outputs, and the aggregated ratings are presented in Figure 3. Across the full question set, ChatVTE achieved significantly higher scores than ChatGPT in both accuracy (mean 4.46, SD 0.16 vs mean 4.11, SD 0.20; $Z=4.287$; $r=0.553$; $P<.001$) and completeness (mean 4.41, SD 0.20 vs mean 3.98, SD 0.25; $Z=4.429$; $r=0.572$; $P<.001$). In the analysis restricted to the 8 items on postdischarge management, ChatVTE again significantly outperformed the comparator for accuracy (mean 4.50, SD 0.13 vs mean 4.00, SD 0.19; $Z=2.546$; $r=0.637$; $P=.01$) and completeness (mean 4.25, SD 0.19 vs mean 4.09, SD 0.13; $Z=2.546$; $r=0.559$; $P=.03$). The consistency rating

was also higher for ChatVTE (mean 4.53, SD 0.16) than for ChatGPT (mean 4.25, SD 0.17; $Z=4.289$; $r=0.554$; $P<.001$). Regarding perceived safety—defined as the likelihood of potentially harmful misinformation—ChatVTE received a lower score compared to ChatGPT (mean 0.09, SD 0.12 vs mean 0.16, SD 0.18; $Z=-2.530$; $r=0.327$; $P=.01$), indicating a safer profile for ChatVTE. For instance, in response to the specific prompt about managing missed anticoagulant doses (question 28), ChatGPT received its highest risk rating (score=2, denoting moderate concern). In contrast, only 1 reviewer assigned ChatVTE a score of 1, indicating minimal concern. The 4 independent raters showed strong concordance, with their ratings falling within a 1-point range for 86.7% (26/30) or more of the items across all domains and models. No item exhibited a rating discrepancy exceeding 2 points on the scale.

Figure 3. Comparison between ChatVTE and ChatGPT for validity evaluation. Scores for accuracy, completeness, and consistency are rated on a scale of 1 to 5, with higher scores indicating better performance. For safety, scores range from 0 to 4, with lower scores reflecting lower risk and thus higher safety. Q: question.



Patient Experience Assessment

All 25 participants returned valid questionnaires (Figure 4). Baseline demographic and clinical characteristics are summarized in Table 1. The average duration of ChatVTE use was 7.6 (SD 3.9; IQR 5.5-9.5) days. Figure 5 shows that the highest mean rating was for response timeliness (item 3:

4.84/5), while the lowest was for emotional support (item 7, 1.92/5). Overall satisfaction and willingness to recommend the platform to other patients with VTE received mean scores above 4.4 on the 5-point scale, indicating generally favorable user acceptance.

Figure 4. Flowchart for evaluating the use of ChatVTE in patients with venous thromboembolism (VTE).

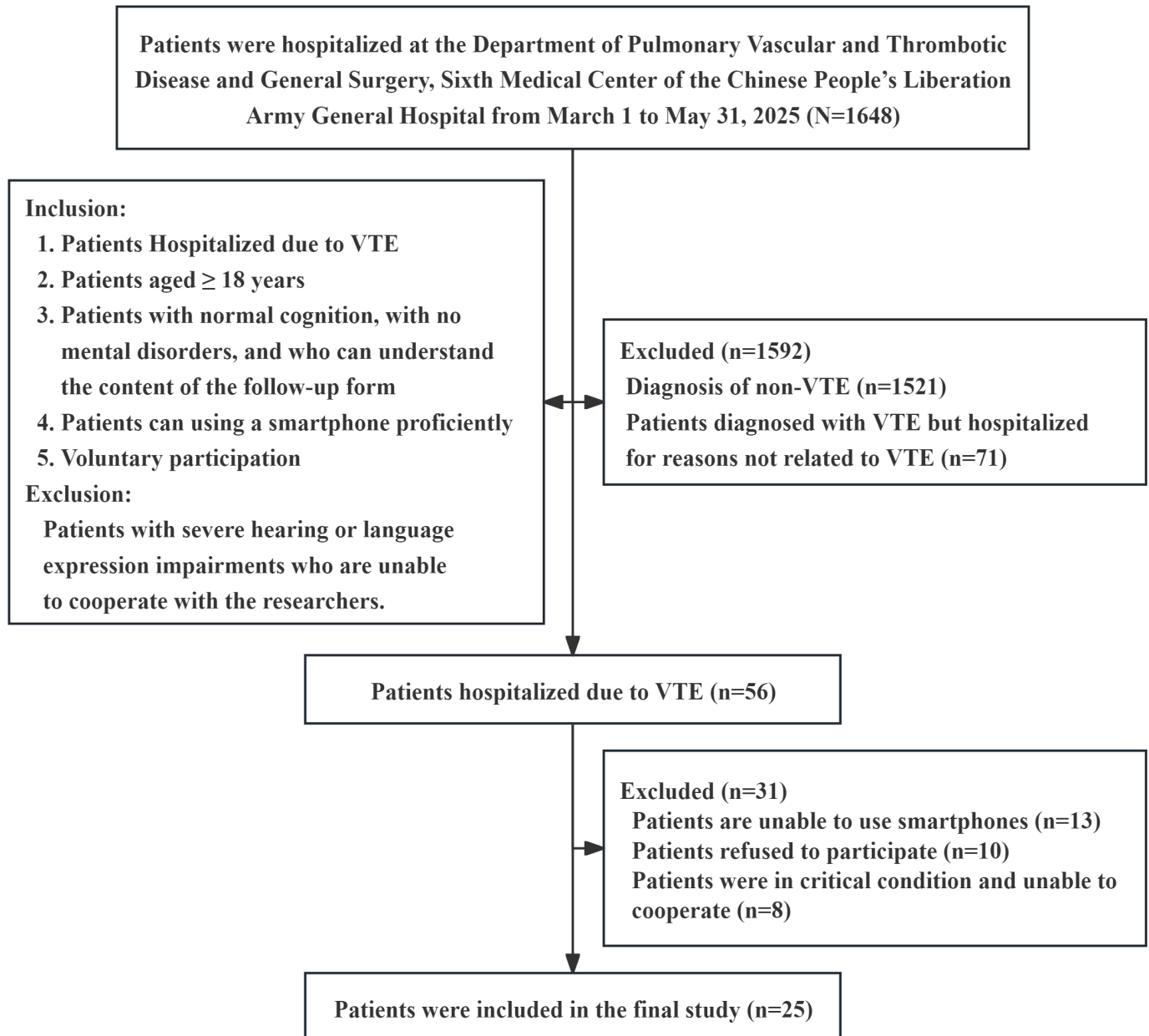


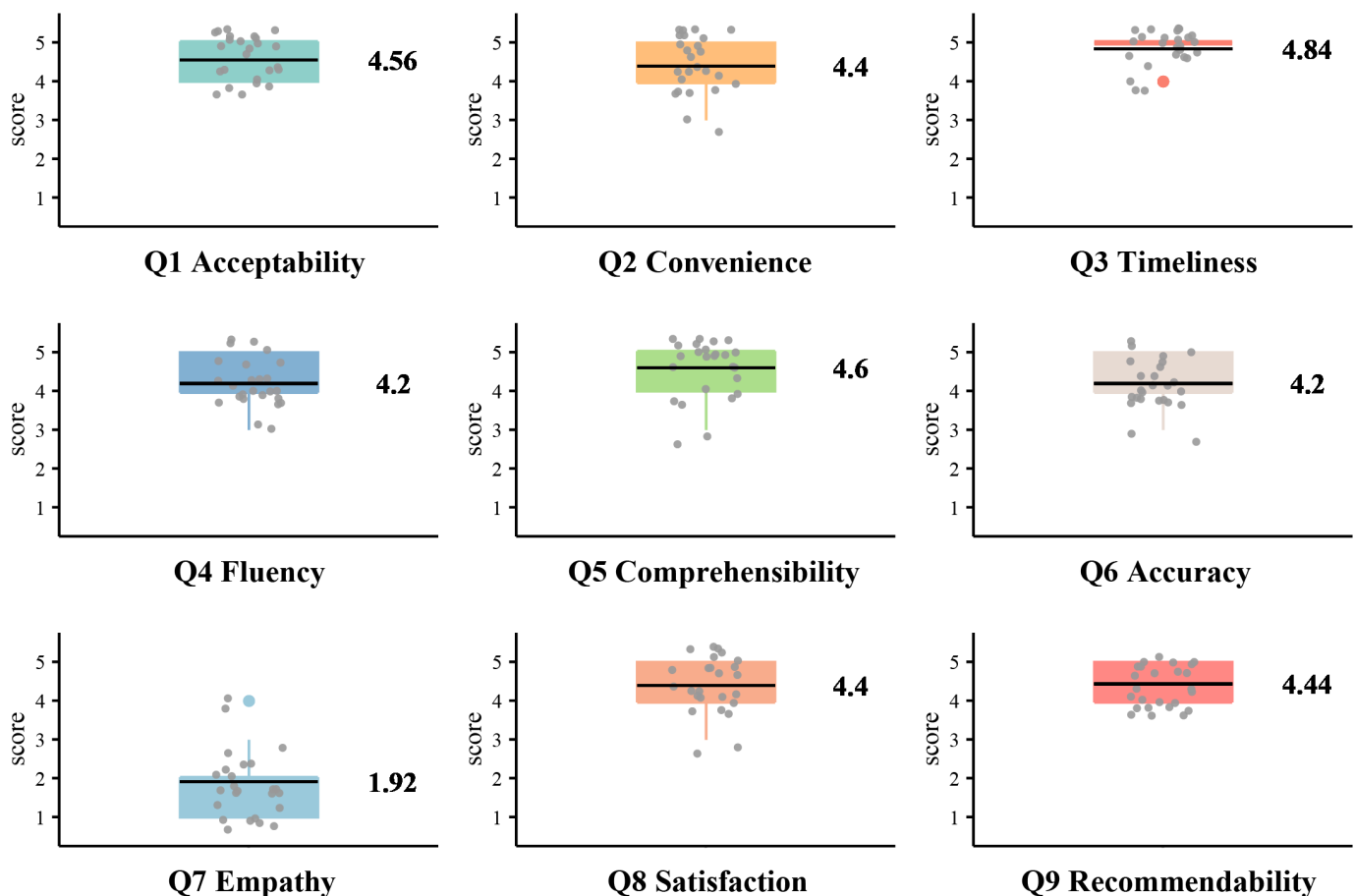
Table 1. Baseline characteristics of all patients (N=25).

Characteristics	Patients
Age (years), mean (SD)	55.4 (13.2)
Age group (years), n (%)	
≥ 60	11 (44)
< 60	14 (56)
Female, n (%)	10 (40)

Characteristics	Patients
Education level, n (%)	
More than high school	16 (64)
Less than high school	9 (36)
BMI (kg/m ²), mean (SD)	25.5 (3.5)
≥28 kg/m ^{2a} , n (%)	6 (24)
Comorbidities, n (%)	19 (76)
Cancer	8 (32)
Diabetes	5 (20)
Hypertension	4 (16)
Rheumatic immune disease	3 (12)
ChatVTE use time (days), mean (SD)	7.6 (3.9)

^aBMI ≥28 kg/m² is defined as obesity according to the Chinese guidelines.

Figure 5. Evaluation results for patients with venous thromboembolism after using ChatVTE in various dimensions (higher scores mean better evaluations). Q: question.



Discussion

Principal Findings

In this study, we introduced ChatVTE, an LLM-based system specialized for VTE care. It innovatively integrates patient health information, personalized patient needs, RAG-augmented LLM, and a digital virtual physician to form a convenient and comprehensive patient education and support platform. The dual-stage evaluation demonstrated ChatVTE's strong performance in expert and patient assessments.

ChatVTE aims to facilitate patient-centered VTE care, which requires precise management based on personal health information. The system uses an LLM-enhanced multimodal OCR with patient-uploaded data, enabling continuous tracking of disease progression. Beyond acute treatment, poor self-management engagement and low behavioral adherence remain critical barriers to effective long-term VTE care. Evidence suggests that interventions based on KAP and HBM can improve patients' self-management, and integrating patients' values enables LLMs to generate content and recommendations that are tailored to patients' preferences [20-23]. To address these barriers, we incorporated the

evaluation of patients' KAP levels and HBM status into ChatVTE. By combining patients' personal health information with their psychosocial characteristics, ChatVTE can deliver customized VTE educational resources, self-management strategies, and health-promoting behaviors, thereby promoting patient-centered care.

RAG addresses inherent shortcomings of general-purpose LLMs, such as hallucination, and its effectiveness is well-established [11-13,24-26]. ChatVTE leverages a RAG-based LLM, which is significantly superior to a general-purpose LLM (ChatGPT). The high accuracy and completeness of the model's response demonstrated its feasibility in the care of patients with VTE. In expert assessments, no responses received a risk score ≥ 2 (moderate or higher). This preliminary finding suggests that the system may contribute to reducing the risk of adverse outcomes caused by misguidance, offering a promising tool for patient education and care. Higher response consistency indicated that it can provide more stable and reliable guidance for patients with VTE, reducing clinical decision-making confusion and patient compliance risks caused by conflicting information. On the basis of a structured questionnaire survey among patients hospitalized for VTE, ChatVTE proved overall highly satisfactory (average score $\geq 4.2/5$), although there remained room for improvement in empathy. Furthermore, ChatVTE used TTS and lip-synch technology to convert generated text into streaming video, thereby creating an interactive digital virtual VTE physician. This significantly enriches the patient's experience and is particularly user-friendly for patients with limited digital health literacy (such as older individuals and those with limited education).

Comparison With Prior Work

Consistent with prior studies, ChatVTE further demonstrates the feasibility of applying a RAG-augmented LLM in specific diseases (VTE) [13,27,28]. Most general-purpose and medical-specific LLMs are text-based, which is relatively monotonous and provides a poor user experience for older patients and those with limited education. An LLM-driven, 3D, hyperrealistic interactive intelligent digital human system named MetaTutor has changed this paradigm [29]. It can interact with users through voice and text and simulate facial expressions and body movements based on the content of the conversation. ChatVTE created an interactive digital virtual VTE physician that was the first digital virtual physician driven by specialty-specific VTE knowledge retrieval LLM. It provides patients with an immersive communication experience, capable of explaining complex medical information and providing accurate responses to queries. As ChatVTE was primarily designed for patient self-service, its interactive capabilities within virtual scenarios are currently less advanced than those of MetaTutor. However, the use of virtual scenarios will be a key area for future optimization of ChatVTE. Regarding response consistency, Kelly et al [30] designed an augmented LLM for patients with type 2 diabetes mellitus. Their evaluation method involved repeatedly inputting identical questions, which does not account for the natural variation in how patients phrase queries in real-world settings. This approach may not fully

reflect the model's adaptability to synonymous and heterogeneous queries. Textual matching based on cosine similarity makes it impossible to judge the consistency of core information transmission in terms of clinical significance. However, this may be related to their original intention of designing the model to improve the health literacy of patients with type 2 diabetes mellitus through robust content output.

ChatVTE's limited empathetic capacity stems from its fundamental lack of genuine emotional experience. Its responses are generated solely by imitating language patterns in training data, which hampers its ability to truly comprehend and contextualize patients' complex emotional states and underlying needs [31]. While optimized for delivering accurate clinical information, ChatVTE's prompts and response templates may insufficiently prioritize the recognition and validation of patient emotions. For instance, the system might correctly answer a factual question about anticoagulation but fail to acknowledge the accompanying anxiety expressed by the patient. However, careful design of LLM prompts (limiting response scope, setting goals and formats, standardizing interaction rules, etc) can improve the empathy of model output results [30,32]. Future versions will integrate emotion-aware prompting to acknowledge patient concerns, incorporate a library of validating response templates (eg, for anticoagulation-related anxiety), and implement a simple keyword-triggered rule to recommend human consultation upon detecting signals of severe distress. These modifications offer practical strategies for mitigating the common issue of insufficient empathy in medical LLMs.

Privacy and security concerns surrounding personal health information make integration into LLMs challenging. Ge et al [33] developed an LLM for liver-related diseases based on a platform that meets protected health information standards, providing new strategies for achieving patient-centered care. Developed for the care and education of patients with VTE, ChatVTE enables the extraction of pertinent health information from protected hospital medical records under strict controls, while also allowing patients to voluntarily upload their clinical information. In addition, the integration of social and psychological status evaluation into patient care has been absent from previous similar studies. The practical application of disease-specific LLM is also important. For instance, NeuroBot, an LLM for neurosurgical patient education, was evaluated using focus groups comprising health care professionals [28]. This subjective method and the exclusive focus on professionals may limit the generalizability of the findings to the target patient population. In contrast, the LLM developed by Adhikary et al [34] for menstrual health education was evaluated more comprehensively among both health care professionals and diverse volunteers.

Future Research Directions

ChatVTE, based on the mVTEA, provides patients with precise and personalized VTE information in a convenient manner. This helps reduce reliance on nonprofessional online content, enhances patients' awareness of disease risks, and improves adherence to preventive measures. Although

it cannot replace clinical consultations, it has the potential to optimize communication between physicians and patients. Future research should consider the following aspects. First, ChatVTE requires further optimization. This includes enhancing its empathetic responses through tailored prompt words and integrating it with the hospital information system. Second, future studies should integrate ChatVTE into the entire process of patient management and evaluate its effectiveness in patient education and achieving patient-centered care in clinical practice, as well as its effects on reducing adverse events and improving long-term prognosis in patients with VTE.

Limitations

This study has several limitations. First, as a single-arm, single-center pilot study with a modest sample size (N=25), the patient experience assessment of ChatVTE limits the generalizability of the findings to other clinical settings and broader patients with VTE. The limited sample size may result in insufficient statistical power to detect anything other than large effects. Consequently, the positive user ratings should be interpreted cautiously as preliminary evidence of acceptability. The lack of a control group in this pilot study means that positive user feedback could be influenced, in part, by the increased attention inherent to participation in a study (eg, the Hawthorne effect) [35]. This is an inherent limitation of a single-arm feasibility study.

Another limitation is the patient experience questionnaire used in this study. While it was refined through a qualitative pilot process to ensure face validity and clinical relevance, it has not yet undergone formal large-scale psychometric validation. Therefore, the findings regarding patient experience should be considered preliminary evidence of user acceptance, and a comprehensive, multicenter evaluation of the ChatVTE system should be conducted in the future.

Acknowledgments

The authors thank all the staff and participants at the Sixth Medical Center of the Chinese People's Liberation Army General Hospital and all the patients who voluntarily took part in the study. They appreciate the DrBreath Company (DrBreath Co, Ltd, Beijing, China) for their technical assistance with the ChatVTE and the mVTEA app.

Funding

The study was funded by the National Key Research and Development Program of China (The Key Project for the Prevention and Treatment of Common and Multiple Diseases; grant 2023YFC2507201).

Data Availability

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Authors' Contributions

BbL contributed to conceptualization, methodology, data curation, investigation, formal analysis, visualization, and writing the original draft. ZgJ contributed to conceptualization, methodology, project administration, validation, writing, reviewing, and editing. ZqZ contributed to the investigation, writing, reviewing, and editing. Hong W contributed to the investigation, writing, reviewing, and editing. Hao W contributed to data curation, validation, writing, reviewing, and editing. HZ contributed to data curation, validation, writing, reviewing, and editing. CL contributed to methodology, software, writing, reviewing, and editing. FQ contributed to methodology, software, writing, reviewing, and editing. YtG contributed to conceptualization, funding acquisition, methodology, project administration, supervision, validation, writing, reviewing, and editing.

Conflicts of Interest

None declared.

In addition, the comparison between ChatVTE and ChatGPT, while offering referential value, is contextually limited. ChatGPT represents a general-purpose LLM without medical-domain optimization, whereas ChatVTE is a proprietary LLM designed for education and management of patients with VTE. Consequently, the observed advantages of ChatVTE should be interpreted primarily as validation of its customized architecture in achieving specific clinical objectives, rather than as a direct, head-to-head comparison of model capabilities. Future evaluations should incorporate medical-specialized LLMs, particularly those focused on VTE, to provide a more comprehensive and equitable assessment. Although the preliminary safety assessment of ChatVTE was encouraging based on expert ratings of a limited question set, real-world deployment must address practical issues such as digital literacy gaps, user trust, data privacy, and ethical compliance. If patients interpret ChatVTE-generated recommendations as definitive medical advice, it could lead to potential harm. Therefore, it is necessary to establish safeguards such as human monitoring and legal disclaimers.

Conclusions

ChatVTE, an augmented LLM-based platform, synthesizes individual clinical information and psychosocial characteristics to enable novel strategies for patient education and patient-centered care. It provides patients with an immersive experience through an interactive artificial intelligence conversational agent. These preliminary findings suggest that ChatVTE may serve as a promising, scalable tool for supporting patient-centered VTE care. Future research should rigorously assess its effectiveness on clinical outcomes, long-term usability, and implementation within diverse health care settings.

Multimedia Appendix 1

Response from ChatVTE to the question “what is venous thromboembolism?”

[[MP4 File \(MP4 video File\), 9763 KB-Multimedia Appendix 1](#)]

Multimedia Appendix 2

The venous thromboembolism-related issues used in the ChatVTE validation process, the corresponding expert scoring criteria, and the structured questionnaire for the patient experience survey.

[[DOCX File \(Microsoft Word File\), 32 KB-Multimedia Appendix 2](#)]

Checklist 1

iCHECK-DH checklist.

[[DOCX File \(Microsoft Word File\), 25 KB-Checklist 1](#)]

References

1. Khan F, Tritschler T, Kahn SR, Rodger MA. Venous thromboembolism. *Lancet*. Jul 3, 2021;398(10294):64-77. [doi: [10.1016/S0140-6736\(20\)32658-1](#)] [Medline: [33984268](#)]
2. Nicolaides AN, Fareed J, Spyropoulos AC, et al. Prevention and management of venous thromboembolism. International consensus statement. Guidelines according to scientific evidence. *Int Angiol*. Feb 2024;43(1):1-222. [doi: [10.23736/S0392-9590.23.05177-5](#)] [Medline: [38421381](#)]
3. Haut ER, Aboagye JK, Shaffer DL, et al. Effect of real-time patient-centered education bundle on administration of venous thromboembolism prevention in hospitalized patients. *JAMA Netw Open*. Nov 2, 2018;1(7):e184741. [doi: [10.1001/jamanetworkopen.2018.4741](#)] [Medline: [30646370](#)]
4. Nana M, Shute C, Williams R, Kokwaro F, Riddick K, Lane H. Multidisciplinary, patient-centred approach to improving compliance with venous thromboembolism (VTE) prophylaxis in a district general hospital. *BMJ Open Qual*. Jul 2020;9(3):e000680. [doi: [10.1136/bmjoc-2019-000680](#)] [Medline: [32718914](#)]
5. Torres MB, Kendall HA, Kerwin A, et al. Venous thromboembolism prevention compliance: a multidisciplinary educational approach utilizing NSQIP best practice guidelines. *Am J Surg*. Nov 2020;220(5):1333-1337. [doi: [10.1016/j.amjsurg.2020.06.059](#)] [Medline: [32709409](#)]
6. Popoola VO, Lau BD, Shihab HM, et al. Patient preferences for receiving education on venous thromboembolism prevention - a survey of stakeholder organizations. *PLoS One*. Mar 31, 2016;11(3):e0152084. [doi: [10.1371/journal.pone.0152084](#)] [Medline: [27031330](#)]
7. Stone TE, Jia Y, Kunaviktikul W. Mobile apps: an effective, inclusive and equitable way of delivering patient and nurse education? *Nurse Educ Today*. Feb 2020;85:104308. [doi: [10.1016/j.nedt.2019.104308](#)] [Medline: [31785570](#)]
8. Aydin S, Karabacak M, Vlachos V, Margetis K. Large language models in patient education: a scoping review of applications in medicine. *Front Med (Lausanne)*. Oct 29, 2024;11:1477898. [doi: [10.3389/fmed.2024.1477898](#)] [Medline: [39534227](#)]
9. Elkin PL, Mehta G, LeHouillier F, et al. Semantic clinical artificial intelligence vs native large language model performance on the USMLE. *JAMA Netw Open*. Apr 1, 2025;8(4):e256359. [doi: [10.1001/jamanetworkopen.2025.6359](#)] [Medline: [40261653](#)]
10. Sharma A, Medapalli T, Alexandrou M, Brilakis E, Prasad A. Exploring the role of ChatGPT in cardiology: a systematic review of the current literature. *Cureus*. Apr 2024;16(4):e58936. [doi: [10.7759/cureus.58936](#)] [Medline: [38800264](#)]
11. Merritt R. What is retrieval-augmented generation, aka RAG? NVIDIA. 2025. URL: <https://blogs.nvidia.com/blog/what-is-retrieval-augmented-generation/> [Accessed 2025-05-26]
12. Li Y, Li Z, Zhang K, Dan R, Jiang S, Zhang Y. ChatDoctor: a medical chat model fine-tuned on a large language model meta-AI (LLaMA) using medical domain knowledge. *Cureus*. Jun 2023;15(6):e40895. [doi: [10.7759/cureus.40895](#)] [Medline: [37492832](#)]
13. Long C, Subburam D, Lowe K, et al. ChatENT: augmented large language model for expert knowledge retrieval in otolaryngology-head and neck surgery. *Otolaryngol Head Neck Surg*. Oct 2024;171(4):1042-1051. [doi: [10.1002/ohn.864](#)] [Medline: [38895862](#)]
14. Liu B, Jin Z, Wang H, Zhang H, Yang Y, Guo Y. Smart technology facilitated patient-centered venous thromboembolism management: pilot study on the digital feasibility. Presented at: ISTH 2024: International Society on Thrombosis and Haemostasis; Jun 22-26, 2024; Bangkok, Thailand. URL: <https://isth2024.eventscribe.net/fsPopup.asp?PresentationID=1432695&mode=presInfo> [Accessed 2026-03-25]
15. Perrin Franck C, Babington-Ashaye A, Dietrich D, et al. iCHECK-DH: guidelines and checklist for the reporting on digital health implementations. *J Med Internet Res*. May 10, 2023;25:e46694. [doi: [10.2196/46694](#)] [Medline: [37163336](#)]

16. Li C, Jin Z, Wang F, Zhang Z, Liu B, Guo Y. A novel QR code-based solution for secure electronic health record transfer in venous thromboembolism home rehabilitation management: algorithm development and validation. *JMIR Rehabil Assist Technol*. Aug 11, 2025;12:e69230. [doi: [10.2196/69230](https://doi.org/10.2196/69230)] [Medline: [40789086](https://pubmed.ncbi.nlm.nih.gov/40789086/)]
17. Viechtbauer W, Smits L, Kotz D, et al. A simple formula for the calculation of sample size in pilot studies. *J Clin Epidemiol*. Nov 2015;68(11):1375-1379. [doi: [10.1016/j.jclinepi.2015.04.014](https://doi.org/10.1016/j.jclinepi.2015.04.014)] [Medline: [26146089](https://pubmed.ncbi.nlm.nih.gov/26146089/)]
18. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Routledge; 1988. ISBN: 9780203771587
19. Fritz CO, Morris PE, Richler JJ. Effect size estimates: current use, calculations, and interpretation. *J Exp Psychol Gen*. Feb 2012;141(1):2-18. [doi: [10.1037/a0024338](https://doi.org/10.1037/a0024338)] [Medline: [21823805](https://pubmed.ncbi.nlm.nih.gov/21823805/)]
20. Mohebbi B, Tol A, Sadeghi R, Mohtarami SF, Shamshiri A. Self-management intervention program based on the Health Belief Model (HBM) among women with gestational diabetes mellitus: a quazi-experimental study. *Arch Iran Med*. Apr 1, 2019;22(4):168-173. [Medline: [31126174](https://pubmed.ncbi.nlm.nih.gov/31126174/)]
21. Wang J, Chen L, Yu M, He J. Impact of knowledge, attitude, and practice (KAP)-based rehabilitation education on the KAP of patients with intervertebral disc herniation. *Ann Palliat Med*. Mar 2020;9(2):388-393. [doi: [10.21037/apm.2020.03.01](https://doi.org/10.21037/apm.2020.03.01)] [Medline: [32233633](https://pubmed.ncbi.nlm.nih.gov/32233633/)]
22. Xu LS, Gao ZG, He M, Yang MD. Effectiveness of the knowledge, attitude, practice intervention model in the management of hypertension in the elderly. *J Clin Hypertens (Greenwich)*. May 2024;26(5):465-473. [doi: [10.1111/jch.14770](https://doi.org/10.1111/jch.14770)] [Medline: [38468407](https://pubmed.ncbi.nlm.nih.gov/38468407/)]
23. Nolan VJ, Balch JA, Baskaran NP, et al. Incorporating patient values in large language model recommendations for surrogate and proxy decisions. *Crit Care Explor*. Aug 1, 2024;6(8):e1131. [doi: [10.1097/CCE.0000000000001131](https://doi.org/10.1097/CCE.0000000000001131)] [Medline: [39132980](https://pubmed.ncbi.nlm.nih.gov/39132980/)]
24. Salihu A, Gadiri MA, Skalidis I, et al. Towards AI-assisted cardiology: a reflection on the performance and limitations of using large language models in clinical decision-making. *EuroIntervention*. Dec 4, 2023;19(10):e798-e801. [doi: [10.4244/EIJ-D-23-00461](https://doi.org/10.4244/EIJ-D-23-00461)] [Medline: [38050992](https://pubmed.ncbi.nlm.nih.gov/38050992/)]
25. Zakka C, Shad R, Chaurasia A, et al. Almanac - retrieval-augmented language models for clinical medicine. *NEJM AI*. Feb 2024;1(2). [doi: [10.1056/aioa2300068](https://doi.org/10.1056/aioa2300068)] [Medline: [38343631](https://pubmed.ncbi.nlm.nih.gov/38343631/)]
26. Miao J, Thongprayoon C, Suppadungsuk S, Garcia Valencia OA, Cheungpasitporn W. Integrating retrieval-augmented generation with large language models in nephrology: advancing practical applications. *Medicina (Kaunas)*. Mar 8, 2024;60(3):445. [doi: [10.3390/medicina60030445](https://doi.org/10.3390/medicina60030445)] [Medline: [38541171](https://pubmed.ncbi.nlm.nih.gov/38541171/)]
27. Chen X, Zhao Z, Zhang W, et al. EyeGPT for patient inquiries and medical education: development and validation of an ophthalmology large language model. *J Med Internet Res*. Dec 11, 2024;26:e60063. [doi: [10.2196/60063](https://doi.org/10.2196/60063)] [Medline: [39661433](https://pubmed.ncbi.nlm.nih.gov/39661433/)]
28. Ho CM, Guan S, Mok PKL, et al. Development and validation of a large language model-powered chatbot for neurosurgery: mixed methods study on enhancing perioperative patient education. *J Med Internet Res*. Jul 15, 2025;27:e74299. [doi: [10.2196/74299](https://doi.org/10.2196/74299)] [Medline: [40663377](https://pubmed.ncbi.nlm.nih.gov/40663377/)]
29. Song Y, Xiong W. Large language model-driven 3D hyper-realistic interactive intelligent digital human system. *Sensors (Basel)*. Mar 17, 2025;25(6):1855. [doi: [10.3390/s25061855](https://doi.org/10.3390/s25061855)] [Medline: [40292977](https://pubmed.ncbi.nlm.nih.gov/40292977/)]
30. Kelly A, Noctor E, Ryan L, van de Ven P. The effectiveness of a custom AI chatbot for type 2 diabetes mellitus health literacy: development and evaluation study. *J Med Internet Res*. May 5, 2025;27:e70131. [doi: [10.2196/70131](https://doi.org/10.2196/70131)] [Medline: [40324160](https://pubmed.ncbi.nlm.nih.gov/40324160/)]
31. Sorin V, Brin D, Barash Y, et al. Large language models and empathy: systematic review. *J Med Internet Res*. Dec 11, 2024;26:e52597. [doi: [10.2196/52597](https://doi.org/10.2196/52597)] [Medline: [39661968](https://pubmed.ncbi.nlm.nih.gov/39661968/)]
32. Koranteng E, Rao A, Flores E, et al. Empathy and equity: key considerations for large language model adoption in health care. *JMIR Med Educ*. Dec 28, 2023;9:e51199. [doi: [10.2196/51199](https://doi.org/10.2196/51199)] [Medline: [38153778](https://pubmed.ncbi.nlm.nih.gov/38153778/)]
33. Ge J, Sun S, Owens J, et al. Development of a liver disease-specific large language model chat interface using retrieval-augmented generation. *Hepatology*. Nov 1, 2024;80(5):1158-1168. [doi: [10.1097/HEP.0000000000000834](https://doi.org/10.1097/HEP.0000000000000834)] [Medline: [38451962](https://pubmed.ncbi.nlm.nih.gov/38451962/)]
34. Adhikary PK, Motiyani I, Oke G, et al. Menstrual health education using a specialized large language model in India: development and evaluation study of MenstLLaMA. *J Med Internet Res*. Jul 16, 2025;27:e71977. [doi: [10.2196/71977](https://doi.org/10.2196/71977)] [Medline: [40669074](https://pubmed.ncbi.nlm.nih.gov/40669074/)]
35. McCambridge J, Witton J, Elbourne DR. Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. *J Clin Epidemiol*. Mar 2014;67(3):267-277. [doi: [10.1016/j.jclinepi.2013.08.015](https://doi.org/10.1016/j.jclinepi.2013.08.015)] [Medline: [24275499](https://pubmed.ncbi.nlm.nih.gov/24275499/)]

Abbreviations

HBM: health belief model

KAP: knowledge, attitude, and practice

LLM: large language model
mVTEA: mobile venous thromboembolism app
OCR : optical character recognition
RAG: retrieval-augmented generation
TTS: text-to-speech
VTE: venous thromboembolism

Edited by Bradley Malin; peer-reviewed by Gayathri Surianarayanan, Maria Chatzimina, Syrowatka Ania; submitted 21.Aug.2025; final revised version received 12.Mar.2026; accepted 13.Mar.2026; published 06.Apr.2026

Please cite as:

Liu Bbin, Jin Zgeng, Zhang Zqi, Wang H, Wang H, Zhang H, Li Czhen, Qi F, Guo Ytao

Immersive, Interactive, Intelligent Patient Educational System for Venous Thromboembolism (ChatVTE): Development and Validation Study

JMIR AI 2026;5:e82775

URL: <https://ai.jmir.org/2026/1/e82775>

doi: [10.2196/82775](https://doi.org/10.2196/82775)

© Bin bin Liu, Zhe geng Jin, Zhe qi Zhang, Hong Wang, Hao Wang, Hui Zhang, Chang zhen Li, Fei Qi, Yu tao Guo. Originally published in JMIR AI (<https://ai.jmir.org>), 06.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR AI, is properly cited. The complete bibliographic information, a link to the original publication on <https://www.ai.jmir.org/>, as well as this copyright and license information must be included.