

Review

Application of AI in Hypertension Health Education: Scoping Review

Haoran Chen, MM; Shenglan Xiao, MM; Tong Wan, MM; Gui Li, MM; Yanhong Peng*, MM; Zhimin Wang*, PhD

School of Nursing and The Second Affiliated Hospital, Hengyang Medical School, University of South China, Hengyang, China

*these authors contributed equally

Corresponding Author:

Zhimin Wang, PhD

School of Nursing and The Second Affiliated Hospital

Hengyang Medical School, University of South China

28 West Changsheng Road

Hengyang

China

Phone: 86 13974733239

Email: 153462814@qq.com

Abstract

Background: Hypertension is a major global health challenge, and effective health education is crucial for improving patients' self-management. Traditional health education approaches are often limited by insufficient personalization, accessibility, and scalability. Artificial intelligence (AI), including natural language processing, machine learning, and large language models (LLMs), offers promising solutions to address these limitations. However, evidence regarding AI applications in hypertension health education has not been comprehensively synthesized.

Objective: This scoping review aimed to summarize the current evidence on AI applications in hypertension health education, and identify research gaps to inform future research and practice.

Methods: This review followed the Joanna Briggs Institute methodology and PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines. Six databases (PubMed, Embase, Web of Science, Cochrane Library, CINAHL, and Scopus) were searched from January 2015 to June 2026. Eligibility criteria were developed using the participant-concept-context framework. Two reviewers independently conducted study screening and data extraction. Study designs were classified using the Mixed Methods Appraisal Tool framework. Consistent with scoping review methodology, no formal quality assessment was performed. Findings were synthesized narratively and presented using evidence gap maps, tables, and figures.

Results: A total of 24 studies from 11 countries were included, comprising 6 randomized controlled trials, 4 nonrandomized trials, 11 quantitative descriptive studies, and 3 mixed methods studies. Most studies were published between 2024 and 2026. In total, 3 AI application scenarios were identified: rule-based health education, data-driven adaptive health education, and generative AI-driven health education. Natural language processing was the most widely applied technology, and LLM-based applications increased rapidly after 2023. However, generative AI studies were predominantly proof-of-concept evaluations and lacked randomized clinical validation. Health education was rarely implemented as a standalone intervention and was typically embedded within multifunctional AI platforms. Outcomes were categorized using the Digital Health Scorecard Framework across 4 domains: technology, clinical, usability, and cost. Technical accuracy and blood pressure outcomes were the most frequently reported measures, whereas no study evaluated economic outcomes.

Conclusions: This first scoping review of AI applications in hypertension health education identified a mismatch between rapid advances in generative AI and the limited availability of rigorous clinical evidence. Three major research gaps were identified: (1) the lack of standardized core outcome sets covering technical, behavioral, clinical, and implementation domains; (2) limited development of hybrid architectures integrating LLM with structured medical knowledge bases; and (3) the absence of evaluation frameworks that satisfy both regulatory and implementation requirements. AI appears most suitable as a complement to, rather than a replacement for, clinician-delivered education. Future research should prioritize rigorous clinical validation, economic evaluation, multicultural adaptation, and health literacy equity to ensure that AI-driven health education reduces rather than exacerbates disparities in hypertension control.

Trial Registration: OSF Registries osf.io/4wv3f/; <https://osf.io/4wv3f/overview>

J Med Internet Res 2026;28:e95596; doi: [10.2196/95596](https://doi.org/10.2196/95596)

Keywords: artificial intelligence; hypertension; health education; large language models; knowledge graph

Introduction

Rationale

Hypertension is a predominant global chronic condition and has become an important risk factor for many diseases [1]. The high and increasing global burden of hypertension presents a major health challenge, as it contributes to morbidity and mortality from cardiovascular and kidney diseases and imposes substantial financial costs on society [2]. Without effective interventions, the prevalence and absolute burden of hypertension will continue to rise [3,4].

The management of hypertension is a major challenge worldwide, with its control influenced by a variety of factors, including poor treatment adherence, inappropriate medication regimens, lifestyle, and socioeconomic status [5]. Some of these factors, such as poor adherence and unhealthy lifestyles, highlight a key issue: patients' inadequate self-management skills [6]. Targeted health education is key to improving self-management skills and benefits patients [7,8]. In addition, numerous studies indicate that health education can be used as a tool to promote adherence to improve blood pressure control in patients with hypertension [9-13]. Health education is a continuous, dynamic, and planned teaching-learning process that spans the entire life cycle. Through an equal partnership between professionals and clients, it empowers individuals to proactively change their lifestyles in order to achieve positive health outcomes [14].

However, limited clinical staff and time resources often make it difficult to implement this ideal model on a large scale in practice, resulting in most health education remaining at the level of traditional, one-way, and brief information dissemination [15,16]. Traditional health education methods, such as brochures and verbal instruction, have several limitations: they are time-consuming, offer standardized content that lacks personalization, and make continuous follow-up difficult [17,18]. Therefore, new strategies and tools are needed, and recent advances in artificial intelligence (AI) present a promising solution to these limitations.

In recent years, AI has been applied across multiple fields of medicine, playing a crucial role in areas such as clinical decision support, medical image analysis, and genomics research [19]. With the advancement of technology, AI has also provided new approaches for the management of hypertension [4]. AI, particularly natural language processing (NLP), machine learning (ML), and large language model (LLM), holds immense potential for understanding patient needs, generating personalized content, and providing dynamic feedback [20,21]. Applying AI to hypertension health education could yield positive outcomes. Although AI has been widely applied in the field of hypertension, there

remains a gap in the literature regarding how AI can enhance hypertension health education.

Several reviews have explored the role of AI in hypertension management and patient education. Aydin et al [22] conducted a scoping review on the application of LLM in patient education within the medical field. A 2025 review further explored the application of LLM in chronic disease management tasks, with patient education accounting for 62% of the included studies [23]. However, it did not specifically focus on hypertension. A recent scoping review in 2026 synthesized 33 studies on the application of LLM in hypertension care, emphasizing model optimization strategies and evaluation methods rather than AI as a health education tool [24]. Overall, while these reviews have explored the clinical utility, technical performance, or general applicability of AI in patient education, none have systematically examined the specific applications of AI technology in hypertension health education. This research gap urgently requires dedicated study.

Objectives

For the reasons outlined earlier, systematically mapping the existing evidence on the use of AI in hypertension health education is warranted. Therefore, this scoping review aims to synthesize the published literature on the application of AI in hypertension health education. The specific objectives are (1) to identify the types of AI technologies used and their application scenarios, (2) to characterize the methodological approaches or research designs adopted in this domain, and (3) to catalog the outcome measures used to evaluate AI-based hypertension health education.

Methods

Overview

In this study, we used the Joanna Briggs Institute's scoping review framework to map the research landscape regarding the application of AI technologies in health education for hypertension [25,26]. To ensure the reliability of our findings and their practical applicability, we conducted a comprehensive search, systematic screening, and structured data extraction. As this is a scoping review, no formal critical appraisal of study quality was performed [27]. This review is reported in accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines [28], and a complete checklist is provided as [Checklist 1](#).

Protocol and Registration

The protocol for this scoping review was registered with the Open Science Framework (OSF registration number: 4wv3f).

Eligibility Criteria

The inclusion criteria for this scoping review were based on the Joanna Briggs Institute Scope Review Methodology

Guide and structured using the participants, concept, context framework. The inclusion and exclusion criteria are presented in [Textbox 1](#).

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- Population: Adults (≥ 18 years) across the full hypertensive disease trajectory, from individuals with high-normal blood pressure at risk of developing hypertension to those with diagnosed and treated hypertension.
- Concept: Studies that used any form of artificial intelligence (AI) technology, defined as the use of computational methods to perform tasks that normally require human intelligence. Eligible AI technologies included, but were not limited to, machine learning, deep learning, natural language processing, large language models, expert systems, knowledge bases, and conversational agents.
- Context: Hypertension health education, defined as the provision of information, knowledge, or skills training, aimed at improving self-management, treatment adherence, medication adherence, lifestyle modification, or disease awareness among individuals across the hypertension spectrum. This encompasses knowledge dissemination, lifestyle guidance, health education material generation, and patient education delivered in any setting.
- Types of evidence sources: Peer-reviewed original research papers with full text available. No restrictions were placed on publication language.
- Study design: Randomized or nonrandomized controlled trials, qualitative and quantitative studies, and mixed methods studies.

Exclusion criteria

- Population: Studies that did not include participants within the hypertension disease spectrum, or studies that did not report relevant findings.
- Concept: Studies in which the intervention or technology did not involve an AI component as defined. For example, traditional web-based educational platforms without AI-driven functionality, standard telemonitoring without intelligent processing, or purely human-delivered education.
- Context: Studies addressing contexts other than hypertension health education or patient education for populations with hypertension. For example, AI applications exclusively for hypertension diagnosis, risk prediction, drug discovery, or clinical decision support without an educational component.
- Study types other than original research: Review, meta-analysis, editorial, commentary, letter, conference abstract, dissertation, book, book chapter, preprint, and protocol.
- Types of evidence sources: Study not subjected to peer review.

Information Sources

We conducted a systematic search of the following 6 electronic databases: PubMed, Embase, Web of Science, Cochrane Library, CINAHL, and Scopus.

Search

The search strategy for this study followed the PRISMA-S guidelines (an extension of the PRISMA [Preferred Reporting Items for Systematic Reviews and Meta-Analyses] statement for reporting literature searches in systematic reviews) [29]. The initial search was conducted in January 2026. Following iterative optimization of the search strategy, the Scopus database was added in May 2026, and a new search was performed. The search strategy was updated again in June 2026, and a second search was conducted to include the latest literature and any studies missed by the previous search strategy. The initial search strategy covered 5 databases: PubMed, Embase, Web of Science, Cochrane Library, and

CINAHL. Detailed search strategies for all databases are provided in [Multimedia Appendix 1](#). In accordance with the guidelines, this study used a 3-step search strategy. The first step involved a preliminary, limited-scope search of PubMed to analyze keywords in the titles and abstracts of relevant papers, as well as the index terms used to describe the papers, thereby developing a comprehensive search strategy. The second step involved a comprehensive search of all 6 databases using all identified keywords and index terms. To ensure reproducibility, the complete electronic search strategy for at least 1 database is provided as a [Multimedia Appendix 1](#). The third step involved a supplementary search, which included a manual review of the reference lists of all included studies to identify other relevant papers not captured by the database searches. The search was limited to papers published between January 2015 and June 2026 to ensure relevance to contemporary technological advancements. No language restrictions were applied to the search. The PubMed search strategy is detailed in [Textbox 2](#).

Textbox 2. PubMed search strategy.

“Hypertension”[Mesh] OR “Blood Pressure”[Mesh] OR hypertens*[tiab] OR “high blood pressure”[tiab] OR “elevated blood pressure”[tiab] OR “raised blood pressure”[tiab] OR “uncontrolled blood pressure”[tiab] OR “blood pressure control”[tiab] OR “blood pressure management”[tiab] OR “BP control”[tiab] OR “BP management”[tiab] OR “essential hypertension”[tiab] OR “primary hypertension”[tiab] OR “uncontrolled hypertension”[tiab] OR “resistant hypertension”[tiab] OR “arterial hypertension”[tiab] OR “hypertensive patient”[tiab] OR “hypertensive individual”[tiab] OR “hypertensive adult”[tiab] OR “systolic hypertension”[tiab] OR “diastolic hypertension”[tiab] OR antihypertens*[tiab] OR “lowering blood pressure”[tiab] OR “high BP”[tiab]

AND

“Artificial Intelligence”[Mesh] OR “Machine Learning”[Mesh] OR “Deep Learning”[Mesh] OR “Natural Language Processing”[Mesh] OR “Neural Networks, Computer”[Mesh] OR “Decision Support Systems, Clinical”[Mesh] OR “Expert Systems”[Mesh] OR “artificial intelligence”[tiab] OR AI[tiab] OR “machine learning”[tiab] OR “deep learning”[tiab] OR “neural network”[tiab] OR “natural language processing”[tiab] OR NLP[tiab] OR “large language model”[tiab] OR LLM[tiab] OR LLMs[tiab] OR “LLM-based”[tiab] OR “LLM-driven”[tiab] OR ChatGPT[tiab] OR GPT[tiab] OR “generative AI”[tiab] OR “generative artificial intelligence”[tiab] OR “generative pretrained transformer”[tiab] OR “expert system”[tiab] OR “knowledge graph”[tiab] OR “knowledge base”[tiab] OR “clinical decision support”[tiab] OR “decision support system”[tiab] OR “chatbot”[tiab] OR “chat-bot”[tiab] OR “conversational agent”[tiab] OR “virtual assistant”[tiab] OR “intelligent system”[tiab] OR “recommender system”[tiab] OR “predictive model”[tiab] OR “prediction model”[tiab] OR “random forest”[tiab] OR “support vector machine”[tiab] OR SVM[tiab] OR “reinforcement learning”[tiab] OR “transformer model”[tiab] OR BERT[tiab] OR “bidirectional encoder”[tiab] OR “retrieval-augmented generation”[tiab] OR RAG[tiab] OR “text mining”[tiab] OR “speech recognition”[tiab] OR “fuzzy logic”[tiab] OR “Bayesian network”[tiab] OR “ontology”[tiab] OR “supervised learning”[tiab] OR “unsupervised learning”[tiab] OR “data mining”[tiab] OR “pattern recognition”[tiab] OR “computational intelligence”[tiab]

AND

“Health Education”[Mesh] OR “Patient Education as Topic”[Mesh] OR “Self Care”[Mesh] OR “Self-Management”[Mesh] OR “Patient Compliance”[Mesh] OR “Health Promotion”[Mesh] OR “Health Communication”[Mesh] OR “Health Literacy”[Mesh] OR “health education”[tiab] OR “patient education”[tiab] OR “health promotion”[tiab] OR “health communication”[tiab] OR “health information”[tiab] OR “patient information”[tiab] OR “patient teaching”[tiab] OR “patient counselling”[tiab] OR “patient counseling”[tiab] OR “self-management”[tiab] OR “self-management education”[tiab] OR “self care”[tiab] OR “self-care”[tiab] OR “lifestyle modification”[tiab] OR “lifestyle intervention”[tiab] OR “lifestyle change”[tiab] OR “behavioral intervention”[tiab] OR “behavioural intervention”[tiab] OR “behavior change”[tiab] OR “behaviour change”[tiab] OR “health coaching”[tiab] OR “medication adherence”[tiab] OR “treatment adherence”[tiab] OR “therapeutic adherence”[tiab] OR “patient adherence”[tiab] OR “patient empowerment”[tiab] OR “health knowledge”[tiab] OR “patient knowledge”[tiab] OR “patient engagement”[tiab] OR “educational intervention”[tiab] OR “educational program”[tiab] OR “educational material”[tiab] OR “health behavior”[tiab] OR “health behaviour”[tiab] OR “dietary advice”[tiab] OR “dietary education”[tiab] OR “exercise counseling”[tiab] OR “exercise education”[tiab] OR “lifestyle guidance”[tiab] OR “lifestyle advice”[tiab] OR “health advice”[tiab] OR “consumer health information”[tiab]

Selection of Sources of Evidence

After the search was completed, all retrieved records were organized and imported into EndNote (Clarivate), and duplicate entries were removed. Two independent reviewers (HC and SX) screened the titles and abstracts to assess whether they met the inclusion criteria. For potentially relevant papers, the same 2 independent reviewers retrieved the full texts and conducted a detailed assessment based on the inclusion criteria. The reasons for excluding full-text papers that did not meet the inclusion criteria were documented and presented in the PRISMA flowchart. In accordance with guidelines for supplementary searches, the reference lists of all included studies were manually searched to identify other relevant literature cited in the included studies. Records identified through these supplementary searches underwent the same screening process described earlier. At each stage of the screening process, any disagreements among reviewers were resolved through discussion or with the assistance

of a third reviewer (ZW). The search results and the study screening process have been fully reported in the PRISMA-ScR flowchart.

Data Charting Process

Two independent reviewers (HC and SX) extracted data from the included studies using a standardized data extraction form developed in Microsoft Excel. The 2 reviewers pilot-tested the form on 3 randomly selected included studies and refined it before proceeding with the full data extraction. Disagreements that arose during data extraction were resolved through discussion or by consulting a third reviewer (ZW). When necessary, the corresponding authors of the included studies were contacted to request missing or supplementary data.

Data Items

The following data items were extracted from each included study:

- Basic information: First author, year of publication, and country.
- Study design and methods: Research design classified via Mixed Methods Appraisal Tool (MMAT) categories, study setting, sample size, and duration.
- AI technical characteristics: AI technology type, specific AI technology or model name, core AI techniques used, application scenario, and brief description of application scenario.
- Characteristics of health education: Health education content or domains.
- Outcome measures: All outcome measures reported by the authors and summary of reported conclusions.

Critical Appraisal of Individual Sources of Evidence

No formal critical appraisal of individual evidence sources was undertaken. As a scoping review, this study was designed to map the breadth and characteristics of the available evidence. Because of the heterogeneity in study designs and the exploratory nature of the field, all eligible primary studies were retained for data charting and narrative synthesis. The absence of formal critical appraisal is acknowledged as a limitation and is taken into consideration when interpreting the findings.

Synthesis of Results

Charted data were synthesized using narrative synthesis and organized by AI technology type, application scenario, health education characteristics, and outcome measures. Findings were summarized using tables, figures, and an evidence gap map.

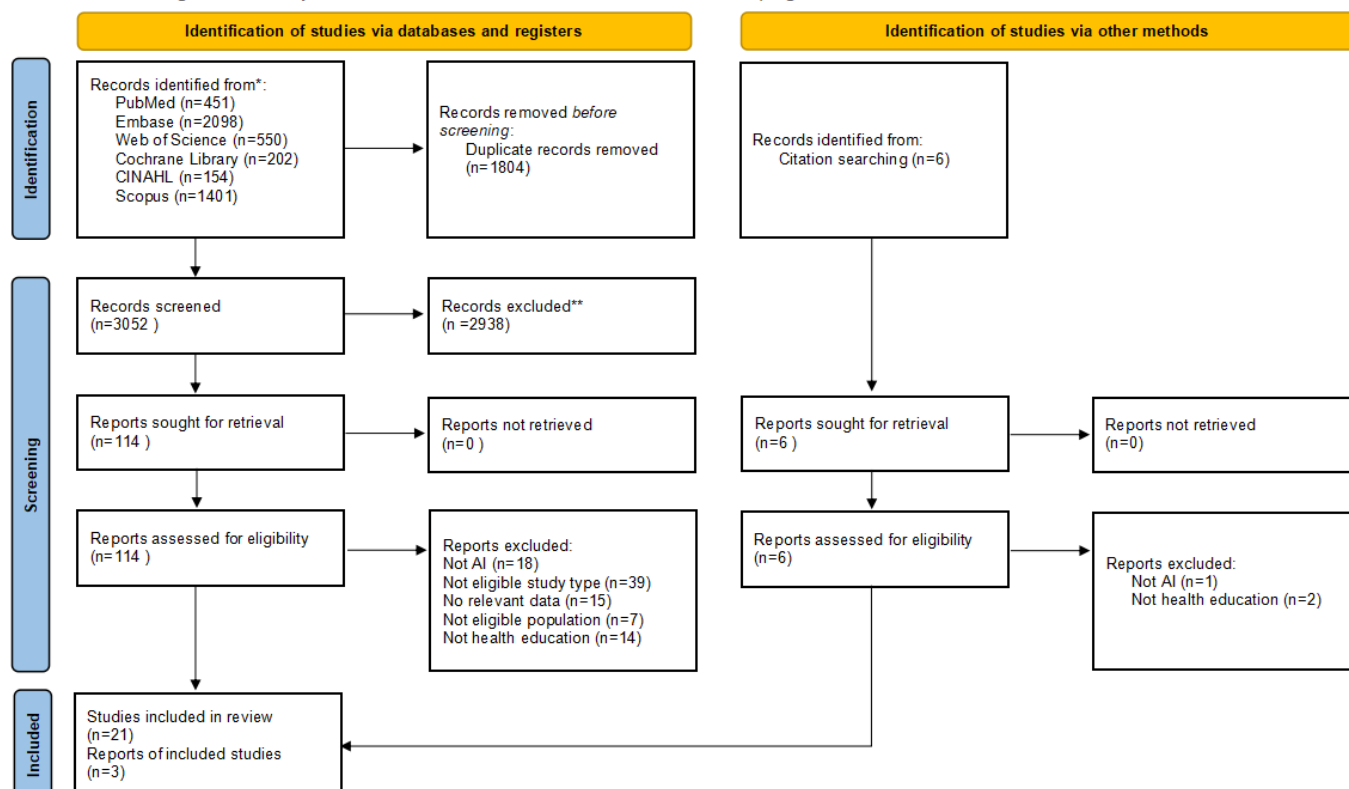
Results

Selection of Sources of Evidence

The study screening process is summarized in the PRISMA flowchart (Figure 1). A total of 4856 records were retrieved from the databases. Among these, 1804 were duplicates, leaving 3052 records to proceed to the screening stage. Of these, 2938 records were excluded because they did not meet the inclusion criteria. A total of 114 studies were downloaded for full-text screening. Following full-text review, 21 studies were included. Through citation searching, an additional 6 records were identified from these studies, of which 3 were ultimately included. This review ultimately included 24 papers sourced from databases and reference lists.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram for study selection process for a scoping review of application of artificial intelligence in hypertension health education (2020-2026) (reproduced from Page et al [30], which is published under Creative Commons Attribution 4.0 International License [31]). AI: artificial intelligence.

PRISMA 2020 flow diagram for new systematic reviews which included searches of databases, registers and other sources



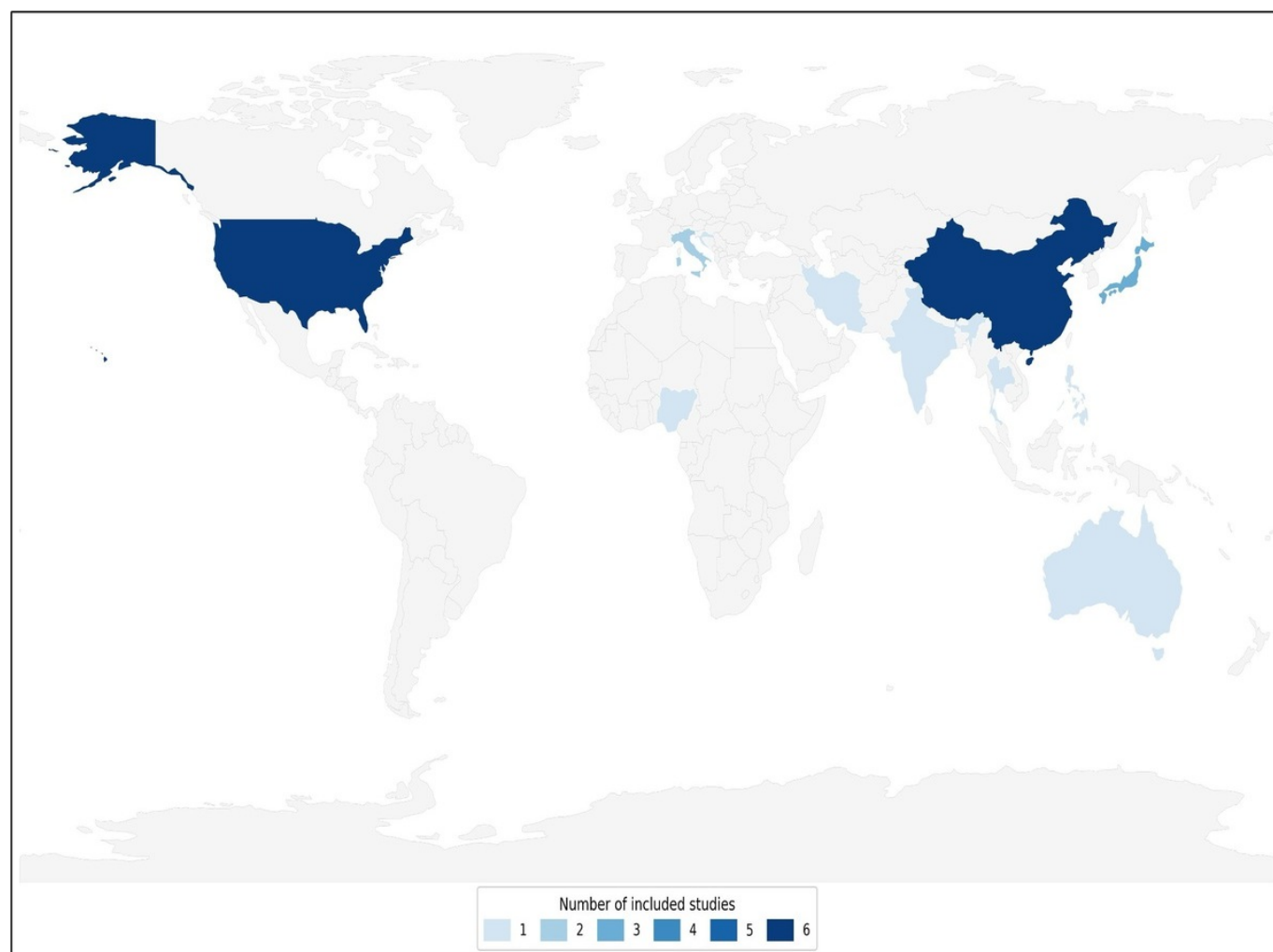
Characteristics of Sources of Evidence

This review included a total of 24 studies published between 2020 and 2026. The majority of these studies (17/24, 71%)

were published between 2024 and 2026, while studies published before 2022 accounted for only 13% (3/24). These studies spanned 11 countries, with China and the United

Nigeria, and Thailand). [Figure 2](#) summarizes the distribution of the 24 included studies by the first author's country of affiliation.

Figure 2. Included studies by the number of publications from the country of the first author.



No formal critical appraisal of individual evidence sources was undertaken. As a scoping review, this study was designed to map the breadth and characteristics of the available evidence.

We classified the study designs of each study based on the MMAT framework [32] and summarized the methodologies used. It should be noted that the MMAT is typically used to assess the quality of mixed methods research. However, in this scoping review, we used only its study design

classification framework to categorize the literature and did not conduct a formal quality assessment. This approach aligns with the methodological principles of scoping reviews, which aim to map the evidence landscape within a specific field rather than critically assess the quality of included studies [25, 33]. Among the 24 included studies, quantitative descriptive studies constituted the largest proportion (n=11, 46%), followed by randomized controlled trials (RCTs; n=6, 25%), quantitative nonrandomized trials (n=4, 17%), and mixed methods studies (n=3, 13%). The main characteristics of the studies included in this review are presented in Table 1. A summary table of characteristics compiled from the data of the included studies can be found in Multimedia Appendix 2 [34-57].

Table 1. The main characteristics of the studies included in author, country, research design, sample size and groups, and duration.

Author (year)	Country	Research design	Sample size and groups	Duration
Persell et al (2020) [34]	United States	Quantitative RCT ^a	• n=333 randomized (intervention=166, control=167) • n=297 completed follow-up	• 6 months

Author (year)	Country	Research design	Sample size and groups	Duration
Griffin et al (2021) [35]	United States	Mixed methods study	n=15 adults with hypertension prescribed ≥ 1 medication	Single time point
Kario et al (2021) [36]	Japan	Quantitative RCT	- n=390 randomized (digital therapeutics=199, control=191) - n=372 completed 12-week follow-up	24 weeks
Gutierrez and Sakulbumrungsil (2023) [37]	Philippines	Quantitative RCT	- n=417 randomized (intervention=214, control=203) - n=401 completed 6-month follow-up	6 months
Griffin et al (2023) [38]	United States	Quantitative descriptive study	n=10 patients with hypertension	Single time point
Sakane et al (2023) [39]	Japan	Quantitative RCT	- n=78 men who are overweight or obese aged 40-69 years with hypertension (intervention=39, control=39) - Retention=95%	12 weeks
Yano et al (2024) [40]	Japan	Quantitative descriptive study	20 questions generated by ChatGPT - Evaluated by 3 blinded certified hypertension or nephrology specialists	Single time point
O'Hagan et al (2023) [41]	Australia	Quantitative descriptive study	15 hypertension-related questions (2 separate question sets with different wording) - Evaluated by 2 independent reviewers on 2 separate occasions	Three evaluation time points: February, April, and May 2023
Almagazzachi et al (2024) [42]	United States	Quantitative descriptive study	100 questions, each posed to ChatGPT 3 times by 3 different study staff - Reviewed by board-certified internal medicine physician	Single evaluation period
Lee et al (2024) [43]	United States	Quantitative descriptive study	52 ACC^b frequently asked questions on hypertension 2 AI^c chatbots (ChatGPT 3.5 and Gemini 1.0) 4 prompt forms (no prompt, patient-friendly, physician-level, and statistics or references) - Total 416 responses analyzed	Single evaluation period
Vinufancis et al (2024) [44]	India	Quantitative descriptive study	10 common patient queries about hypertension 2 AI chatbots (ChatGPT and ChatSonic) 2 internal medicine physician evaluators - Responses assessed per query	Single evaluation period
Leitner et al (2024) [45]	United States	Quantitative nonrandomized trial	- n=141 adults with hypertension - n=128 at 12 weeks - n=102 at 24 weeks	24 weeks
Niko et al (2024) [46]	Iran	Quantitative descriptive study	10 HBPM^d questions, answered twice per chatbot - Evaluated by 3 certified cardiologists	Single evaluation period
Sun et al (2024) [47]	China	Quantitative RCT	- n=68 enrolled, randomized 1:1 - n=54 analyzed (experimental=23, control=31)	12 weeks
Xu et al (2024) [48]	China	Mixed methods study	5 patients with hypertension comorbidities 24 multidisciplinary experts evaluated	Single evaluation period (December 2023 to February 2024)
Aguzzi et al (2025) [49]	Italy	Quantitative descriptive study	21 QA^e test pairs (from augmented dataset of 1473 records) - Multiple RAG^f strategies compared, domain expert evaluation	Single evaluation period
Antia et al (2025) [50]	Nigeria	Quantitative nonrandomized trial	- n=50 adults on hypertension treatment - n=48 completed follow-up 2 (4%) lost to follow-up	30 days

Author (year)	Country	Research design	Sample size and groups	Duration
Jelic et al (2025) [51]	Croatia	Mixed methods study	- n=5136 users of Megi (an AI-based digital health tool for hypertension management) - n=125 users analyzed from online survey of Megi	24 months (for quantitative data) - Single survey time point (for qualitative data)
Montagna et al (2025) [52]	Italy	Quantitative descriptive study	- Simulated patient query dataset 128 questions for BERT^g score 210 responses evaluated by domain experts	Single evaluation period
Moolsart and Kritpolviman (2025) [53]	Thailand	Quantitative nonrandomized trial	n=91 older adults aged 60-75 years with uncontrolled hypertension (experimental=46, comparison=45)	12 weeks
Wang et al (2025) [54]	China	Quantitative nonrandomized trial	- n=51 patients with hypertension from 3 centers - n=6 clinicians 10 FAQ^h for BPⁱ Coach comparison	30 days
Wang et al (2026) [55]	China	Quantitative descriptive study	50 hypertension-related questions across 4 domains 8 model configurations (4 base+4 HEART^j-enhanced) 3 clinicians blind evaluation	Single evaluation period
Wang et al (2026) [56]	China	Quantitative descriptive study	- Internal testing: BKQA^k (n=20), BPQA^l (n=200), RFQA^m (n=139), DMQAⁿ (n=200) - External testing: n=107 suspected patients with hypertension - Benchmarked against 3 physicians	- Internal: August 2024 to February 2025 - External: February 2025 to June 2025
Yao et al (2026) [57]	China	Quantitative RCT	- n=62 patients with hypertension (intervention=31, control=31) - CPET^o subsample n=24	8 weeks

^aRCT: randomized controlled trial.

^bACC: American College of Cardiology

^cAI: artificial intelligence.

^dHBPM: home blood pressure monitoring.

^eQA: question answering.

^fRAG: retrieval-augmented generation.

^gBERT: Bidirectional Encoder Representation from Transformers.

^hFAQ: frequently asked question.

ⁱBP: blood pressure.

^jHEART: Hypertension Enhancing Answer Retrieval Tool.

^kBKQA: blood pressure knowledge question and answer.

^lBPQA: blood pressure question and answer.

^mRFQA: risk factor question and answer.

ⁿDMQA: decision-making question and answer.

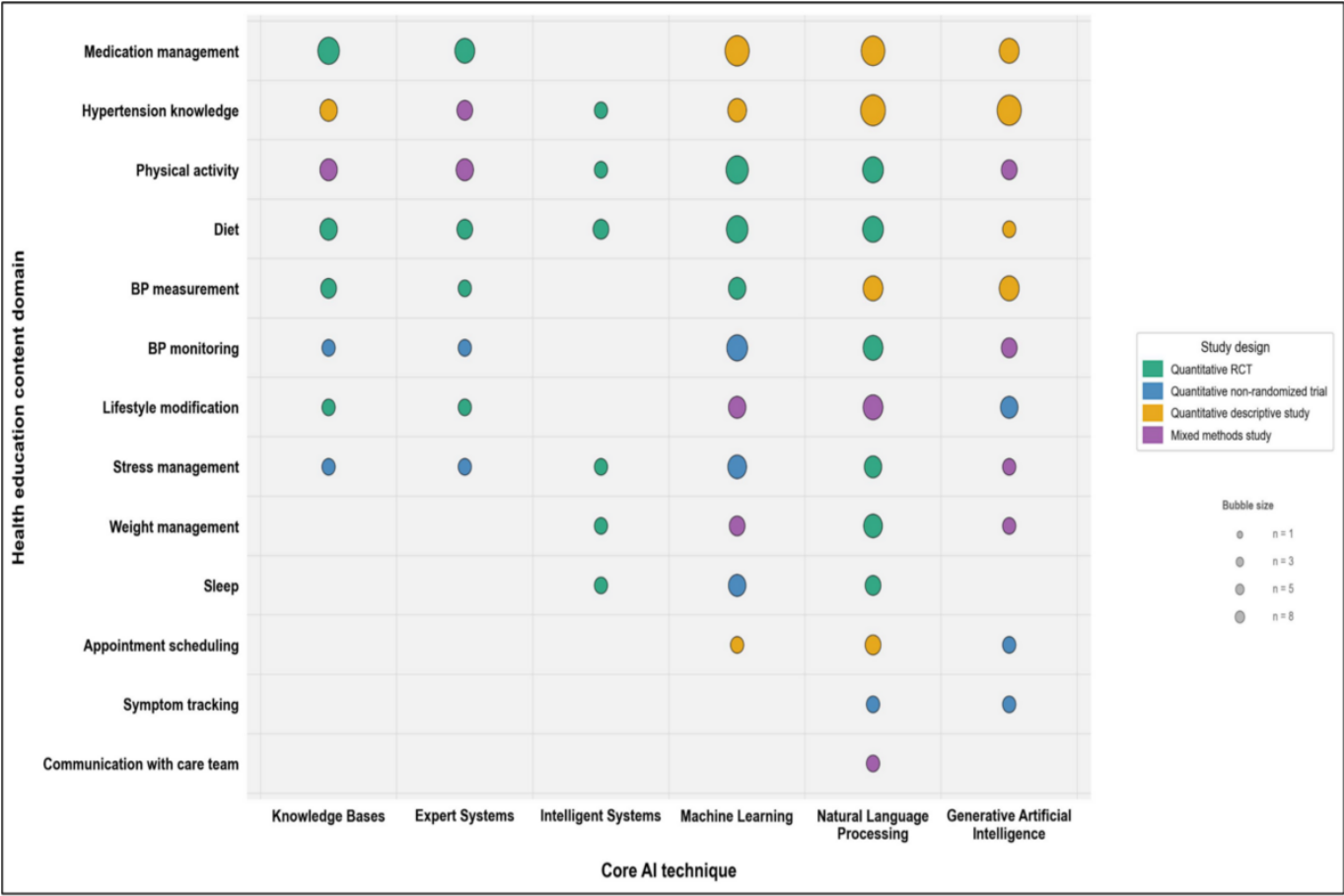
^oCPET: cardiopulmonary exercise testing.

Synthesis of Results

Overview

In this study, we analyzed data from the 24 included studies and created an evidence gap map (Figure 3) to illustrate the core AI technologies used in each study.

Figure 3. Evidence gap map of core AI technologies used in hypertension health education across the 24 included studies. AI: artificial intelligence; BP: blood pressure; RCT: randomized controlled trial.



Among the included studies, NLP was the most widely used technology; nearly all studies involved text or dialogue processing for knowledge transfer, interactive learning, and behavioral guidance in hypertension health education. For example, rule-based dialogue agents and virtual nurse chatbots are typical applications of NLP technology, providing patients with personalized education and self-management support [34,36].

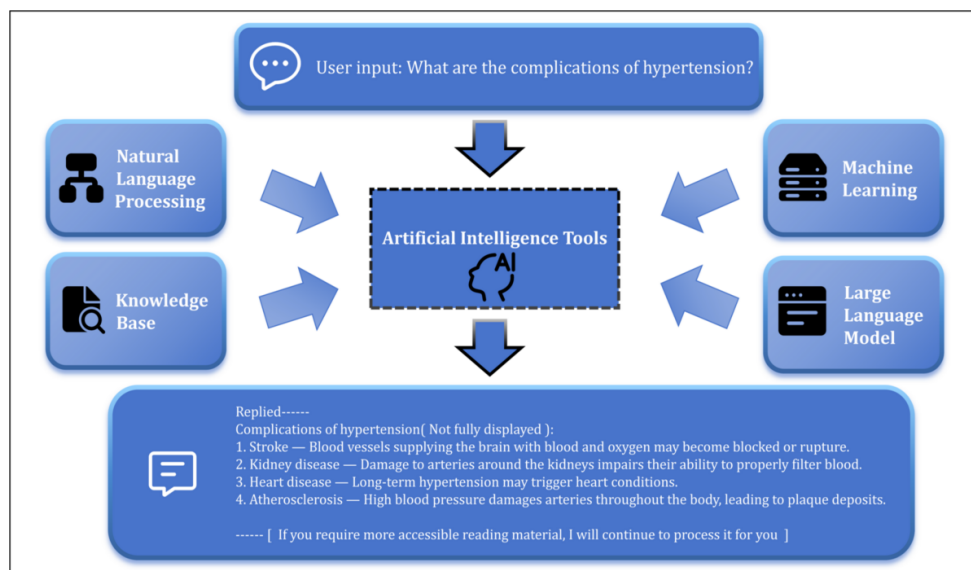
The introduction of generative AI marks a new development in this field, with its applications increasing significantly since 2023. In particular, the emergence of LLMs, such as the study by O’Hagan et al [41], which first evaluated the application of ChatGPT in hypertension health education, has led to a rapid increase in the number of LLM-related studies since 2024. This indicates that generative AI holds potential for improving educational interactions, personalizing information, and generating content.

Regarding the distribution of technology categories, expert system and knowledge base remain prevalent, indicating that

traditional AI methods still hold value in the management of structured educational content and knowledge transfer. Furthermore, NLP and ML technologies are frequently applied together in many studies, suggesting that text processing and algorithm-driven intelligent assistance are jointly supporting the implementation of hypertension health education.

The evidence gap map reveals that, despite the widespread application of NLP and ML technologies, the use of generative AI is still in its infancy. The integration of specific AI technologies with certain areas of health education, such as symptom tracking, clinician-patient communication, and appointment management, remains relatively scarce, providing a clear direction for future research. Overall, AI applications in the field of hypertension health education are rapidly evolving from rule-based and traditional algorithms toward generative and interactive technologies.

As shown in Figure 4, we demonstrate how AI technology is applied to health education for hypertension.

Figure 4. A conceptual model for an artificial intelligence–based hypertension health education system.

Application Scenario of AI

Based on data extracted from 24 studies, we have categorized the applications of AI in hypertension health education into three types (Tables 2 and 3): (1) rule-based health education, (2) data-driven adaptive health education, and (3) generative AI–driven health education. Among these, rule-based health education and generative AI–driven health education are the most common. Rule-based health education is characterized by content determined by predefined rules and knowledge structures, thereby offering determinism and interpretability. Data-driven adaptive health education uses ML models trained on user behavior and physiological data to provide personalized intervention recommendations or

real-time feedback to individuals while adaptively adjusting educational content. Generative AI–driven health education leverages LLMs, combined with techniques, such as retrieval-augmented generation (RAG) and intelligent agents, to support open-ended question-and-answering and personalized dialogue, thereby delivering health education. It is important to distinguish between studies that evaluated LLM outputs against reference standards (proof-of-concept accuracy assessments) and those that deployed LLM-based tools in interactive educational settings with real patients; the majority of generative AI studies in this review fall into the former category, reflecting the early developmental stage of this application scenario.

Table 2. Classification and comparative characteristics of AI application scenarios in hypertension health education.

Application scenario	Technical foundation	Content determination method	Adaptability	Interaction
Rule-based health education	ES ^a , KB ^b , decision trees	Predefined rules and knowledge structures	Low, fixed paths	Push of fixed graphics or texts, structured courses
Data-driven adaptive health education	Traditional ML ^c (eg, collaborative filtering and reinforcement learning)	User behavior and physiological data models	Medium, dynamically adjusts with data	Personalized recommendations, real-time feedback reminders
Generative AI ^d -driven health education	LLM ^e +RAG ^f or agent	Generative dynamic synthesis	High, open-domain Q&A ^g	Multiturn dialogue, natural language Q&A

^aES: expert system.

^bKB: knowledge base.

^cML: machine learning.

^dAI: artificial intelligence.

^eLLM: large language model.

^fRAG: retrieval-augmented generation.

^gQ&A: question and answer.

Table 3. Distribution of the 24 included studies by artificial intelligence (AI) application scenario.

Application scenario	Studies	Year range	Study design
Rule-based health education	[34-39,47,51]	2020-2025	- Quantitative RCT^a (n=4) - Mixed methods study (n=1) - Quantitative descriptive study (n=2)

Application scenario	Studies	Year range	Study design
Data-driven adaptive health education	[45,53,57]	2024-2026	- Quantitative nonrandomized trial (n=1) - Quantitative nonrandomized trial (n=2)
Generative AI-driven health education	[40-44,46,48-50,52,54-56]	2023-2026	- Quantitative RCT (n=1) - Quantitative descriptive study (n=10) - Quantitative nonrandomized trial (n=1) - Mixed methods study (n=2)

^aRCT: randomized controlled trial.

These 24 studies were categorized into 3 distinct application scenarios and exhibited clear patterns in terms of both time span and research methods. Rule-based systems (n=8) spanned the entire review period (2020-2025) and accounted for 4 of the 6 RCTs. Data-driven adaptive interventions (n=3) began to emerge in 2024, primarily using nonrandomized designs. Generative AI-driven applications (n=13) have dominated the literature since 2023 but remain concentrated

in the proof-of-concept phase. In total, 8 of the 11 studies used quantitative descriptive designs, and none were evaluated via RCTs.

Characteristics of Health Education

We summarized the characteristics of health education based on data from the included studies, as shown in Table 4.

Table 4. Characteristics of health education: author, health education content or domains, and its role in study.

Author (year)	Health education content or domains	The role of health education in study
Persell et al (2020) [34]	- BP^a self-monitoring - Medication adherence - Weight loss - Diet - Physical activity - Sleep - Stress management	As a component integrated with home blood pressure monitoring (HBPM) and medication tracking to enable interventions
Griffin et al (2021) [35]	- Medication management - Medication refills - Communication with care team - Healthy lifestyles (diet, exercise, and BP tracking)	Providing health education content to support self-management of hypertension
Kario et al (2021) [36]	- Salt restriction - Body weight control - Regular exercise - Improving sleep condition - Stress coping - Reducing alcohol intake - BP management knowledge	As a part of a digital therapeutics system
Gutierrez and Sakulbumrungsil (2023) [37]	- Medication knowledge (reason for taking, side effects, duration, and monitoring) - BP monitoring; lifestyle modifications - What to do if dose missed - Proper medication storage	As a part of a pharmacist-led expert system
Griffin et al (2023) [38]	- Medication information (list, reminders, and side effects) - Medication refills - Healthy recipes - Appointment scheduling - BP tracking and sharing with providers	Providing health education content to support self-management of hypertension
Sakane et al (2023) [39]	- Healthy eating habits - Exercise habits - Lifestyle habits - BP monitoring - Daily steps (5000/7000/8000/10,000 targets) - Weight management	As a part of a smartphone weight-loss app
Yano et al (2024) [40]	All kinds of hypertension information produced due to the inquiries	Large language models provide health education information

Author (year)	Health education content or domains	The role of health education in study
O'Hagan et al (2023) [41]	All kinds of hypertension information produced due to the questions	Large language models provide health education information
Almagazzachi et al (2024) [42]	All kinds of hypertension information produced due to the questions	Large language models provide health education information
Lee et al (2024) [43]	All kinds of hypertension information produced due to the questions	Large language models provide health education information
Vinufrancis et al (2024) [44]	All kinds of hypertension information produced due to the questions	Large language models provide health education information
Leitner et al (2024) [45]	- Physical activity (steps and active time) - Sleep hygiene - Stress management - Dietary choices (salt, alcohol, red meat, and fruits or vegetables) - Medication adherence - BP self-monitoring	As a part of an artificial intelligence (AI)-powered lifestyle coaching program
Niko et al (2024) [46]	- Self-preparation before measurement (clothing, rest, and avoidance of stimuli) - Body position and cuff use - Number of measurement repetitions - Correct recording and reading of HBPM	Large language models provide health education information
Sun et al (2024) [47]	- Exercise prescriptions (frequency, intensity, type, time, volume, and progression) - DASH^b diet (types and quantities of food) - Medication adherence - BP monitoring frequency and technique	As part of a smart health promotion system based on the WeChat platform
Xu et al (2024) [48]	- Exercise prescription (expected health benefits, exercise principles—FITT^c, weekly plan, movement guidance, and precautions) - Disease-specific education (hypertension, diabetes, COPD^d, Parkinson, gout, and chronic nephritis)	Large language models provide health education information
Aguzzi et al (2025) [49]	All kinds of hypertension knowledge: BP monitoring, lifestyle modifications, medication management	Large language models provide health education information
Antia et al (2025) [50]	- Hypertension education (disease knowledge and risk factors) - Medication adherence reminders - Clinic appointment reminders - Symptom tracking - Lifestyle modification guidance	As a part of a WhatsApp-based generative AI chatbot
Jelic et al (2025) [51]	- BP monitoring - Medication adherence - Physical activity - Weight management - Lifestyle modification - Stress reduction	As a part of a chatbot based on a large language model
Montagna et al (2025) [52]	Hypertension general knowledge	Large language models provide health education information
Moolsart and Kritpolviman (2025) [53]	- BP self-monitoring skills - Dietary management (food tracking with carbohydrate or protein or fat or sugar or sodium details) - Physical activity (step counting and exercise) - Stress management (relaxation breathing and positive thinking) - Medication adherence	As a part of an AI-based self-health monitoring program

Author (year)	Health education content or domains	The role of health education in study
Wang et al (2025) [54]	- BP assessment - Behavior change (lifestyle habits, exercise, diet, sleep, and mental health) - Digital phenotyping education - Medication management; treatment reminders	As a part of a multimodal digital platform for hypertension management
Wang et al (2026) [55]	- BP knowledge (systolic or diastolic, normal values, diagnosis, classification, risk factors, and complications) - BP measurement (methods, cuff selection, timing, frequency, and correct technique) - BP management (diet, fruits or vegetables, alcohol, weight, exercise, sauna, stress, sleep, and e-cigarettes) - Medication management (necessity, tolerance, organ damage, timing, and drug interactions)	Large language models provide health education information
Wang et al (2026) [56]	Hypertension knowledge	Large language models provide health education information
Yao et al (2026) [57]	- Exercise prescription (warm-up, cardiorespiratory endurance training, strength resistance training, balance training, flexibility training, and cooldown) - Exercise safety and posture correction - Lifestyle guidance	As a part of an AI-assisted CPET ^e exercise prescription tool

^aBP: blood pressure.

^bDASH: dietary approaches to stop hypertension.

^cFITT: Frequency, Intensity, Time, Type.

^dCOPD: chronic obstructive pulmonary disease.

^eCPET: cardiopulmonary exercise testing.

Outcomes Measured to Evaluate AI-Based Hypertension Health Education

After analyzing the included studies, we found that their primary research objectives could be broadly categorized into 2 types: one aimed to evaluate the clinical efficacy of AI-based hypertension health education interventions, while the other assessed the performance of LLMs or AI-based

hypertension health education systems. Studies in the first category reported outcomes related to clinically relevant measures. The second category of studies involved the measurement of common computer-related metrics, such as system usability, the accuracy of health education information, and readability. The outcome characteristics of the included studies are shown in [Table 5](#).

Table 5. Outcome measures of included studies.

Author (year)	Outcome measures
Persell et al (2020) [34]	- Primary: SBP^a at 6 months - Secondary: Self-reported antihypertensive medication adherence, home monitoring and self-management practices, measures of self-efficacy associated with blood pressure (BP), weight, and self-reported health behaviors
Griffin et al (2021) [35]	Information-need themes: - Perceptions toward chatbots - Perceived frequency of use - Barriers and facilitators of using
Kario et al (2021) [36]	- Primary: Mean change in 24-hour ambulatory SBP from baseline to 12 weeks - Secondary: Home SBP or DBP^b (morning and evening), office SBP or DBP, heart rate, salt intake, body weight, BMI, app engagement rate, self-reported executive scores for app-guided behaviors, and adverse events
Gutierrez and Sakulbumrungsil (2023) [37]	- Primary: Medication Possession Ratio - Secondary: Medication adherence, SBP, DBP, controlled BP, number of antihypertensive agents, medication changes, and self-perceived knowledge score
Griffin et al (2023) [38]	- Effectiveness: Task completion rate, user error rate, and system error rate - Efficiency: Number of clicks, utterances, and duration of interaction per task

Author (year)	Outcome measures
Sakane et al (2023) [39]	• Satisfaction: System Usability Scale • Qualitative feedback (strengths and shortcomings) • Body weight, BMI, SBP, DBP, adherence to daily self-weighing, pedometer use, BP monitoring, self-reported health behaviors (exercise habits, eating habits, lifestyle habits, and daily steps), and personality traits
Yano et al (2024) [40]	• Appropriateness of responses • Comparison of Japanese versus English response quality (accuracy, comprehensiveness, professionalism, and level of detail) • Gwet agreement coefficient for interrater reliability
O'Hagan et al (2023) [41]	• Readability • Credibility • Accuracy
Almagazzachi et al (2024) [42]	• Accuracy • Reproducibility
Lee et al (2024) [43]	• Proportions of correct, partially correct, and incorrect responses • Flesch-Kincaid grade level • Word count
Vinufrancis et al (2024) [44]	• Global Quality Scale scores • Modified DISCERN scores • Cohen κ for interrater agreement
Leitner et al (2024) [45]	• Primary: SBP and DBP change from baseline to 12 and 24 weeks and percentage change in BP categories (controlled, stage-1, and stage-2) • Secondary: Participant engagement rate and number of manual clinician outreaches
Niko et al (2024) [46]	• Accuracy • Completeness • Reproducibility
Sun et al (2024) [47]	• Primary: SBP, DBP, exercise time, medication adherence, DASH^s adherence, and BP monitoring frequency • Secondary: Weight, SEVR^d, baPWV^e, heart rate, learning performance, diet types or quantity, and weekly adherence curves
Xu et al (2024) [48]	• Accuracy, comprehensiveness, and applicability • Qualitative narrative recommendations from 24 multidisciplinary experts • Kendall concordance coefficients for expert agreement
Aguzzi et al (2025) [49]	• Faithfulness • Medical faithfulness • Domain expert evaluation scores • Chi-square tests with Cramér V for RAG^f versus non-RAG comparison
Antia et al (2025) [50]	• Feasibility: Training time (minutes), proportion able to use bot within 5 minutes • Acceptability: frequency of chats, Chatbot Usability Questionnaire, and Self-made Healthy Heart Assistant Satisfaction Questionnaire • Preliminary efficacy: SBP or DBP, hypertension knowledge test, and medication adherence
Jelic et al (2025) [51]	• User retention at 3, 6, 9, 12, and 24 months • SBP reduction • Spearman correlation between SBP drop and duration of use • Online survey: Usefulness in improving self-management, satisfaction, willingness to use, behavioral response, and unmet needs
Montagna et al (2025) [52]	• Intent recognition: Precision, recall, and accuracy per class and overall • Data extraction accuracy (measure, quantity, format, and overall) • BERT^g score (precision, recall, and F_1 vs GPT-3.5 Turbo reference) • Domain expert evaluation • Chatbot response time
Moolsart and Kripolviman (2025) [53]	• Primary: Hypertension-controlling behavior • Secondary: mean arterial pressure, SBP, and DBP
Wang et al (2025) [54]	• Patient: MAUQ^h subscales—ease of use and satisfaction, system information arrangement, and usefulness • Clinician: Doctor's Software Satisfaction Questionnaire and patient management time • BP coach: Utility, conciseness, completeness, and clarity
Wang et al (2026) [55]	• Accuracy, completeness, consistency, robustness, and security

Author (year)	Outcome measures
	• Overall quality • Interrater reliability • Cohen <i>d</i> effect sizes • Between-group comparisons (base vs HEARTⁱ-enhanced)
Wang et al (2026) [56]	• Internal testing: BP classification accuracy, risk factor stratification accuracy, clinical decision appropriateness, education comprehensiveness and accuracy, discriminative performance (AUC^j), decision curve analysis, IDI^k, and NRI^l • External validation: BP classification accuracy, risk stratification accuracy, clinical decision accuracy, patient-perceived understandability, credibility, and emotional support
Yao et al (2026) [57]	• Primary: 6-minute walk distance • Secondary: SBP, DBP, CPET^m indices (peak VO₂ [mL/min/kg], peak VO₂%pred [%], total exercise time, anaerobic threshold, RERⁿ, maximum load, and heart rate at rest per peak), IPAQ^o, SF-12^p, PHQ-9^q, GAD-7^r, exercise self-efficacy, body weight, handgrip strength (right or left), and waist or hip circumference

^aSBP: systolic blood pressure.
^bDBP: diastolic blood pressure.
^cDASH: dietary approaches to stop hypertension.
^dSEVR: subendocardial viability ratio.
^ebaPWV: brachial-ankle pulse wave velocity.
^fRAG: retrieval-augmented generation.
^gBERT: Bidirectional Encoder Representation from Transformers.
^hMAUQ: mHealth App Usability Questionnaire.
ⁱHEART: Hypertension Enhancing Answer Retrieval Tool.
^jAUC: area under the curve.
^kIDI: integrated discrimination improvement.
^lNRI: Net Reclassification Index.
^mCPET: cardiopulmonary exercise testing.
ⁿRER: respiratory exchange ratio.
^oIPAQ: International Physical Activity Questionnaire.
^pSF-12: 12-item Short Form Health Survey.
^qPHQ-9: Patient Health Questionnaire-9.
^rGAD-7: Generalized Anxiety Disorder-7.

We categorized all studies according to the definitions in the Digital Health Scorecard Framework [58]. The Digital Health Scorecard Framework encompasses 4 domains: technical, clinical, usability, and cost. In this framework, technical refers to evaluating whether a digital health solution can accurately and precisely deliver its claimed functionality, including considerations such as security, interoperability, and system architecture. The clinical dimension focuses on rigorous evaluation of evidence to validate whether the solution has demonstrated capacity to improve specific health outcomes, requiring comparison against relevant clinical gold standards; usability concerns the ease with which users can accomplish intended tasks with minimal effort, encompassing aspects like effectiveness, learnability, and user satisfaction; and cost includes user access fees, technology life-cycle investments, and integration expenses within clinical workflows.

Technical

In the included studies, accuracy was the most frequently evaluated metric. The performance of AI-based hypertension health education systems needs to be compared against reference standards, such as clinical guidelines, medical textbooks, and expert-developed question sets [40-43,46,48,49,52,55,56]. The completeness and comprehensiveness of health education content are also key metrics [46,48,55,

56]. In addition, readability and credibility were evaluated [41], both of which are related to the accuracy and reliability of information delivery. One study further assessed readability using objective measures including the Flesch-Kincaid grade level and response length, providing additional evidence regarding the accessibility of AI-generated educational content [43].

The system’s performance in the face of external disturbances is reflected in the assessment of robustness when handling unexpected inputs [55]. According to the framework, privacy and security are explicitly listed as part of the technical evaluation, which aligns with the current requirement that AI applications in medicine must protect patient data [55]. Other relevant technical metrics include fidelity and medical fidelity [49], intent recognition and data extraction accuracy [52], applicability [48], and the appropriateness and professionalism of responses. Vinufrancis et al [44] additionally evaluated information quality and reliability using the Global Quality Scale and the modified DISCERN instrument, highlighting the importance of assessing the educational value and trustworthiness of AI-generated health information. Domain expert evaluations [49,52] and interrater reliability [40,44,55] serve as complementary methods for validating technical performance.

Clinical

Blood pressure, as a key clinical end point, has been reported in numerous studies, with measurement methods including office blood pressure, home blood pressure, and 24-hour ambulatory blood pressure monitoring [34,36,37,39,45,47,50,51,53,57]. This underscores the need to compare any AI system used in clinical practice against established clinical gold standards. Other cardiovascular-related parameters, such as heart rate [36,47], subendocardial viability ratio, and brachial-ankle pulse wave velocity [47], as well as cardiopulmonary exercise test parameters and 6-minute walk distance [57], were also included in the assessment of clinical outcomes.

The process indicators emphasized by this framework are particularly evident in the assessment of medication adherence, specifically treatment adherence or adherence to clinical guidelines. The indicators used include medication possession rates, self-reported adherence, and the Morisky Medication Adherence Scale [34,37,47,50]. The number of antihypertensive medications and medication adjustments [37] are also used as process-related indicators. In one study [53], blood pressure control behaviors served as the primary outcome measure, and behavioral monitoring was reported—including blood pressure monitoring frequency, daily self-weighing, pedometer use, and dietary approaches to stop hypertension diet adherence [39,47]. Self-reported health behaviors [34,39] and salt intake [36] also meet the definition of process measures.

Other relevant clinical indicators include self-efficacy related to blood pressure [34] and physical activity [57], as well as knowledge about hypertension [37,47,50], academic performance [47], weight, BMI [34,36,39,47,57], waist, hip circumference [57], grip strength [57], mental health scales [57], and app-guided behavioral adherence scores [36].

Cost

Among the included studies, cost assessment was the least well-developed dimension, which is consistent with the framework's observation that comprehensive cost estimation is a complex process. None of the studies included in this review reported formal cost-effectiveness analyses, technology lifecycle costs, or the long-term economic impacts of clinical efficacy improvements.

Usability

The included studies used various methods to assess usability. Griffin et al [38] used the System Usability Scale to capture users' subjective evaluations of ease of use and likability. Task-level efficiency was measured by task completion rates, user error rates, number of clicks, number of voice commands, and interaction duration per task [38]. Chatbot response times [52] and qualitative feedback regarding their strengths and weaknesses were used to assess user experience. Two additional studies also assessed user experience using the "Chatbot Usability Questionnaire" and a custom-designed satisfaction questionnaire [50], combined with online surveys covering dimensions such as practicality,

satisfaction, willingness to use, behavioral responses, and unmet needs [51].

Patient engagement is also a key indicator of usability, with metrics including user retention rates over a 24-month period [51], chat frequency [50], and app engagement rates [36,45]. For example, training time and the proportion of users able to operate the chatbot within 5 minutes [50] are also used to assess usability. Explorations of chatbot cognition, perceived use frequency, and barriers and facilitators [35] provide insights into user needs, reflecting the necessity of user-centered design for AI systems.

Usability for clinicians was assessed using the Physician Software Satisfaction Questionnaire and patient management time [54], indicating that AI-based hypertension health education systems should not increase the burden on clinical staff. The patient-oriented Mobile Health App Usability Questionnaire covered ease of use, satisfaction, system information layout, and practicality [54], while also incorporating patient-perceived comprehensibility, credibility, and emotional support [56], thereby reflecting usability at the subjective level. The evaluation of BP Coach in terms of practicality, simplicity, completeness, and clarity [54] also covered various aspects of effective and practical design.

Discussion

Summary of Evidence

Overview

This scoping review synthesizes the existing evidence on the application of AI in health education for hypertension. A total of 24 studies met the inclusion criteria, and this scoping review yielded 3 key findings.

First, NLP and ML form the technological foundation of AI in hypertension health education. The study identified 3 application scenarios: rule-based health education, data-driven adaptive health education, and generative AI-driven health education. The evolutionary trajectory from rule-based to generative AI methods reflects the overall trend in the application of AI to hypertension health education, with research based on LLM surging since 2023. However, the use of generative AI in hypertension health education is still limited to the proof-of-concept phase. While some studies have used quantitative descriptive designs, none of them have undergone clinical validation through RCT.

Second, health education is generally embedded within multicomponent AI platforms rather than implemented as a standalone intervention; in all 6 RCTs, the educational module was provided concurrently with monitoring, reminder, or clinical decision support functions, making it impossible to assess the specific effects of the educational intervention in isolation.

Third, the included studies used a multidimensional evaluation framework covering the 4 domains of the Digital Health Scorecard, but significant asymmetry was observed. Technical metrics (accuracy, completeness, and readability)

were reported most frequently, clinical outcome metrics (blood pressure and medication adherence) were reported in more than half of the studies, usability metrics (satisfaction and engagement) were reported less than the clinical outcome metrics, and cost assessments were rarely reported. This evaluation gradient, ranging from robust technical validation to a complete lack of economic analysis, indicates that the current evidence base is insufficient to support real-world implementation decisions. This is due to the absence of large-scale RCTs, short follow-up periods (mostly ≤ 12 weeks), and a general lack of cost-effectiveness data. These 3 points collectively constitute the 3 most urgent gaps in the current evidence base that need to be addressed.

Trends in the Application of AI in Hypertension Health Education

Rule-based health education systems, exemplified by early platforms such as the coaching app [34] and the digital therapeutic system [36], represent how AI technology was initially applied to deliver health education. Their strengths lie in content certainty and clinical interpretability: every piece of educational information can be traced back to predefined rules or validated knowledge structures, making these systems inherently traceable and suitable for integration into clinical workflows where transparency is paramount [37]. However, this certainty comes at the expense of flexibility; such systems cannot address novel queries outside their programmed knowledge domain and have limited capacity for personalization beyond predefined hierarchical criteria. This limitation echoes a longstanding critique in health education research that standardized, noncustomized materials often fail to meet the specific needs and circumstances of individual patients [59].

Data-driven adaptive approaches use ML to analyze user behavior and physiological data to enable dynamic educational functions. The personalized coaching platform developed by Leitner et al [45] is a prime example of this approach. By training personalized models, the platform identifies individual-specific associations between lifestyle factors and blood pressure. By providing dynamic feedback and adaptive health recommendations, these systems address the lack of personalization inherent in traditional rule-based frameworks. However, they also face their own challenges: achieving reliable personalization requires vast amounts of individual-level data, which raises concerns about data privacy. Furthermore, the effectiveness of these systems depends largely on users' consistent use of monitoring devices, which may limit their applicability in settings where digital literacy or access to devices is limited [50].

The most transformative development has been the emergence of generative AI and LLM. Recent literature reviews on chronic disease management indicate that the use of LLMs in patient education has experienced explosive growth [22,23], a trend consistent with our findings: studies based on LLM have increased rapidly since 2023. By supporting multiturn natural language dialogue, these models overcome the one-way limitations of traditional health education, thereby enabling patients to actively seek

information rather than passively receive it. However, LLMs also pose significant challenges to the reliability of educational content. As noted in evaluations of medical AI, LLMs remain prone to generating medical claims that appear plausible but lack evidence [60,61]. Within this review, LLM chatbot evaluations for hypertension education have demonstrated generally acceptable factual accuracy, yet revealed persistent concerns regarding readability and expert assessment consistency [42-44]. Patients, however, lack sufficient discernment, a factor that is particularly critical for the implementation of hypertension health education, which requires precise guidance on medication use and lifestyle adjustments [62,63]. To address this, recent studies advocate for the use of hybrid architectures to improve this situation by anchoring LLM to structured medical knowledge bases [64-66]. This aligns with the findings of this review, as multiple studies have used this approach to reduce hallucinations and enhance the accuracy of educational content [49,55,56]. Emerging evidence also suggests that performance gains can be achieved through prompt optimization alone. Li et al [67] demonstrated that structured prompt engineering strategies, particularly guidance-based and self-consistency approaches, substantially improved the accuracy and guideline adherence of LLM in hypertension treatment decision-making. This finding further highlights the potential of combining model architecture, knowledge integration, and prompt design to enhance the reliability and safety of AI-generated hypertension education.

The Outcome Measures Used to Evaluate the Effectiveness of AI Are Heterogeneous

This review found that classifying outcome metrics reveals heterogeneity in current evaluation practices. The technical dimension has consistently received significant attention, with its accuracy, completeness, readability, and reliability typically validated based on clinical guidelines or expert judgment [40-44,46]. This focus aligns with digital health validation pathways, which regard technical reliability as a prerequisite for clinical deployment [58]. However, a significant portion of the literature on LLM technologies for health education in this review remains at the proof-of-concept stage, assessing whether models can generate accurate information, rather than whether the information provided can change patient behavior or improve health outcomes. Currently, clinical validation of the use of LLM for hypertension health education is still in its early stages.

Blood pressure control is the most frequently studied clinical end point; however, most studies have relatively short follow-up periods, and only a few have used randomized designs to validate hard clinical end points based on established hypertension criteria [36,37,47]. Furthermore, the AI technologies used in these studies rarely involve LLM.

Even more striking is the complete absence of cost assessments across all 24 included studies. This omission aligns with evidence from implementation science, which indicates that even technically effective and user-friendly interventions struggle to achieve scale without data on implementation costs, technology lifecycle investments,

and long-term health economic impacts [68,69]. Although usability evaluations covered user satisfaction, engagement rates, and system usability scales [38,50,54], these assessments primarily reflect short-term acceptance rather than long-term behavioral maintenance. Furthermore, research has found that content generated by LLM far exceeds the recommended reading level for patient education materials [41,43], raising fundamental equity concerns: if populations with lower health literacy, who are often older, less educated, and from lower socioeconomic backgrounds [70], cannot access AI-powered health education, these tools may exacerbate rather than alleviate existing inequalities in hypertension control.

Health Education as an Embedded Intervention Component

The AI systems included in this study share a notable characteristic that warrants special attention: health education rarely exists as a standalone intervention. Instead, it is embedded within multifunctional AI-driven platforms as one of their components [36,47,54]. This embedded nature has significant implications for effect attribution. When a multidimensional AI intervention significantly reduces blood pressure, it is impossible to distinguish the specific contribution of the educational component from the effects of the monitoring, reminder, or decision-support components. This challenge reflects a long-standing methodological debate regarding the evaluation of complex interventions, namely, that synergistic effects between components may exceed the sum of their individual contributions [71].

Multicultural Adaptation

This review found that the geographic distribution of the included studies exhibited distinct characteristics, with a significant concentration of research in China and the United States. This may reflect the policy momentum and technological investment in these 2 countries regarding the use of digital tools to manage hypertension. However, this concentration also raises an important question: whether the study findings are generalizable across different health care systems, reimbursement models, and sociocultural contexts. In total, 5 studies were conducted in low- and middle-income countries, 2 of which used relatively simple AI technologies rather than LLM. The study from Nigeria [50] demonstrated that a WhatsApp-based generative AI chatbot not only achieved high user satisfaction but also improved medication adherence in a resource-limited cardiology clinic. This study provides a key proof of concept but also highlights the significant gap in relevant evidence in low- and middle-income countries.

Research on multicultural and multilingual adaptability is almost entirely lacking. Only 2 studies have examined language-related performance differences, with one finding that blinded hypertension specialists rated English responses more favorably than Japanese responses [40], and this performance difference may align with the fact that current LLMs are primarily trained on English-based data. Beyond linguistic factors, although self-management behaviors for

hypertension, including dietary choices, physical activity patterns, medication attitudes, and health care-seeking behaviors, are deeply influenced by cultural context [72,73], studies are needed to explore deeper dimensions of cultural adaptation.

Privacy and Security Considerations

This review found that only a few studies have systematically explored this topic. Montagna et al [52] proposed and compared various privacy-preserving architectures, revealing the inherent trade-offs between performance and data protection. Building on this work, Aguzzi et al [49] extended research on on-device deployment using enhanced RAG mini-language models, pointing the way toward the development of privacy-preserving AI tools that ensure data security without sacrificing performance. The design of digital health interventions must involve the active participation of the target population and be user-centered [74,75]. However, our findings indicate that this requirement has not yet been systematically implemented in AI-driven hypertension health education.

Limitations

We must acknowledge that this scoping review has several limitations. First, we did not conduct a formal quality assessment or risk-of-bias assessment of the included studies. Although this approach is methodologically appropriate for scoping reviews aimed at mapping the evidence landscape rather than synthesizing effect sizes, it means that the strength of evidence from individual studies cannot be quantified. Therefore, our findings should be interpreted with caution. Second, our inclusion criteria were limited to peer-reviewed original research papers for which full-text access was available, thereby excluding preprints, conference abstracts, theses, and gray literature. This exclusion is particularly consequential for AI research, where technological advances are often first described in preprint repositories such as arXiv and medRxiv months or years before formal journal publication. Given the extremely short iteration cycles of AI technology, with major model updates occurring on the order of weeks to months, this limitation may systematically exclude the latest exploratory studies, negative results, and technical evaluations, introducing a temporal lag bias that could underestimate the current scope of AI applications in hypertension health education. Third, although our search strategy did not impose language restrictions, all 24 included studies were published in English-language journals, which may have systematically excluded relevant studies from non-English-speaking countries. Fourth, the included studies exhibited high clinical and methodological heterogeneity in terms of AI technology types, application scenarios, and outcome measures, making direct cross-study comparisons and quantitative meta-analyses impossible. Fifth, although categorizing application scenarios into 3 groups is conceptually useful, certain systems may possess features that span multiple categories. Sixth, the follow-up periods in most studies were relatively short, with only a few exceeding 12 weeks; therefore, evidence regarding the long-term maintenance of healthy behaviors, the sustainability of blood

pressure control, and the reduction in cardiovascular events remains limited [76].

Conclusions

This study used a scoping review methodology to examine the application of AI technologies in health education for hypertension. Most current reviews focus on broader topics, such as chronic disease management and the application of LLMs in hypertension care. Research findings indicate that AI technology plays a significant role in delivering health education on hypertension. In clinical practice, AI should be used as a tool to enhance, rather than replace, the health education provided by clinicians. The nature of educational components embedded within multifunctional AI platforms means that clinicians and developers must design systems in which educational features are purpose-built and assessable, rather than incidental. For the research community, the immediate priority is to bridge the gap between generative AI innovation and rigorous clinical validation. Future trials should prioritize hybrid architectures, combining the conversational flexibility of LLM with the reliability of structured medical knowledge bases via a RAG framework and adopt core outcome sets that encompass technical

accuracy, behavioral change, and cardiovascular end points. For policymakers and health system planners, the widespread lack of cost-effectiveness data poses a fundamental obstacle to resource allocation decisions. Without evidence regarding implementation costs, technology lifecycle investments, and long-term health economic impacts, even technically superior interventions cannot be responsibly scaled up. Equity must be central to future efforts. Furthermore, research has primarily focused on high-income countries, with a near-total lack of studies on multicultural and multilingual adaptability. Combined with evidence suggesting that content generated by LLM exceeds the reading level recommended for patient education, these factors collectively increase the risk that AI-driven hypertension health education may exacerbate rather than narrow existing disparities in hypertension control. To address these interrelated challenges, standardized evaluation frameworks should be developed, cost-effectiveness benchmarks established, accessibility ensured for populations with varying levels of health literacy, and interventions validated across diverse sociocultural and resource settings, all to bring about substantial improvements in self-management for the hundreds of millions of people with hypertension worldwide.

Acknowledgments

The authors declare the use of generative artificial intelligence (GAI) in the research and writing process. According to the GAIDeT (Generative Artificial Intelligence Delegation Taxonomy; 2025) guidelines, the following task was delegated to GAI tools under full human supervision: language polishing and copyediting. The GAI tool used was DeepSeek (Hangzhou DeepSeek Artificial Intelligence Basic Technology Research Co, Ltd). Responsibility for the final manuscript lies entirely with the authors. GAI tools are not listed as authors and do not bear responsibility for the final outcomes. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Funding

This study was supported by the Hunan Province Natural Science Foundation of China (2024JJ9380), the 2024 Scientific Research Project of the Hunan Nursing Association (HNKY202406), and the Emergency Project for COVID-19 Prevention and Control of University of South China (nk20200334).

Authors' Contributions

Conceptualization: HC, ZW

Data curation: TW, GL

Data extraction: HC, SX

Formal analysis: HC, SX

Funding acquisition: ZW

Investigation: HC, SX, TW, GL

Methodology: HC, YP

Project administration: YP

Resources: ZW

Supervision: ZW

Visualization: YP

Writing—original draft: HC

Writing—review and editing: ZW

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search strategy.

[\[DOCX File \(Microsoft Word File\), 30 KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

Characteristics of included studies.

[[XLSX File \(Microsoft Excel File\)](#), 29 KB-[Multimedia Appendix 2](#)]

Checklist 1

PRISMA-ScR checklist.

[[PDF File \(Adobe File\)](#), 584 KB-[Checklist 1](#)]

References

1. Writing Committee Members, Jones DW, Ferdinand KC, et al. 2025 AHA/ACC/AANP/AAPA/ABC/ACCP/ACPM/AGS/AMA/ASPC/NMA/PCNA/SGIM Guideline for the prevention, detection, evaluation and management of high blood pressure in adults: a Report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. *Hypertension*. Oct 2025;82(10):e212-e316. [doi: [10.1161/HYP.000000000000249](#)] [Medline: [40811516](#)]
2. Mills KT, Bundy JD, Kelly TN, et al. Global disparities of hypertension prevalence and control: a systematic analysis of population-based studies from 90 countries. *Circulation*. Aug 9, 2016;134(6):441-450. [doi: [10.1161/CIRCULATIONAHA.115.018912](#)] [Medline: [27502908](#)]
3. Mills KT, Stefanescu A, He J. The global epidemiology of hypertension. *Nat Rev Nephrol*. Apr 2020;16(4):223-237. [doi: [10.1038/s41581-019-0244-2](#)] [Medline: [32024986](#)]
4. Zhou B, Perel P, Mensah GA, Ezzati M. Global epidemiology, health burden and effective interventions for elevated blood pressure and hypertension. *Nat Rev Cardiol*. Nov 2021;18(11):785-802. [doi: [10.1038/s41569-021-00559-8](#)] [Medline: [34050340](#)]
5. Mohl JT, Moreno CA, Sadik K, Singhal M, Rooney A, Ciemins EL. Evaluation of blood pressure control, medication adherence, and therapeutic inertia in US patients with hypertension prescribed multiple antihypertensives. *J Am Heart Assoc*. Jun 17, 2025;14(12):e034787. [doi: [10.1161/JAHA.124.034787](#)] [Medline: [40519194](#)]
6. Konlan KD, Shin J. Determinants of self-care and home-based management of hypertension: an integrative review. *Glob Heart*. 2023;18(1):16. [doi: [10.5334/gh.1190](#)] [Medline: [36968303](#)]
7. Warsi A, Wang PS, LaValley MP, Avorn J, Solomon DH. Self-management education programs in chronic disease: a systematic review and methodological critique of the literature. *Arch Intern Med*. 2004;164(15):1641-1649. [doi: [10.1001/archinte.164.15.1641](#)] [Medline: [15302634](#)]
8. Chodosh J, Morton SC, Mojica W, et al. Meta-analysis: chronic disease self-management programs for older adults. *Ann Intern Med*. Sep 20, 2005;143(6):427-438. [doi: [10.7326/0003-4819-143-6-200509200-00007](#)] [Medline: [16172441](#)]
9. Wang Y, Ye C, Kong L, et al. Independent associations of education, intelligence, and cognition with hypertension and the mediating effects of cardiometabolic risk factors: a Mendelian randomization study. *Hypertension*. Jan 2023;80(1):192-203. [doi: [10.1161/HYPERTENSIONAHA.122.20286](#)] [Medline: [36353998](#)]
10. Soltani D, Azizi B, Behnoush AH, et al. Is lifestyle modification with individual face-to-face education and counseling more effective than usual care for controlling hypertension? A systematic review and meta-analysis of randomized controlled trials. *Health Educ Res*. Sep 20, 2023;38(5):490-512. [doi: [10.1093/her/cyad028](#)] [Medline: [37450326](#)]
11. Chen Y, Li X, Jing G, et al. Health education interventions for older adults with hypertension: a systematic review and meta-analysis. *Public Health Nurs*. May 2020;37(3):461-469. [doi: [10.1111/phn.12698](#)] [Medline: [31943315](#)]
12. Choudhry NK, Kronish IM, Vongpatanasin W, et al. Medication adherence and blood pressure control: a scientific statement from the American Heart Association. *Hypertension*. Jan 2022;79(1):e1-e14. [doi: [10.1161/HYP.000000000000203](#)] [Medline: [34615363](#)]
13. Yusuf S, Joseph P, Rangarajan S, et al. Modifiable risk factors, cardiovascular disease, and mortality in 155 722 individuals from 21 high-income, middle-income, and low-income countries (PURE): a prospective cohort study. *Lancet*. Mar 7, 2020;395(10226):795-808. [doi: [10.1016/S0140-6736\(19\)32008-2](#)] [Medline: [31492503](#)]
14. Pueyo-Garrigues M, Whitehead D, Pardavila-Belio MI, Canga-Armayor A, Pueyo-Garrigues S, Canga-Armayor N. Health education: a Rogerian concept analysis. *Int J Nurs Stud*. Jun 2019;94:131-138. [doi: [10.1016/j.ijnurstu.2019.03.005](#)] [Medline: [30951988](#)]
15. Wang S, Liu K, Tang S, Wang G, Qi Y, Chen Q. Barriers and facilitators to patient education provided by nurses: a mixed-method systematic review. *J Clin Nurs*. Jul 2024;33(7):2427-2437. [doi: [10.1111/jocn.17111](#)] [Medline: [38476038](#)]
16. Hayek M, Ghoul I, Abdullah A, et al. Barriers and facilitators to patient education from nursing perspectives in West bank hospitals: a cross-sectional study. *BMC Nurs*. Jul 1, 2025;24(1):741. [doi: [10.1186/s12912-025-03434-w](#)] [Medline: [40598492](#)]
17. Zwarenstein M, Grimshaw JM, Presseau J, et al. Printed educational messages fail to increase use of thiazides as first-line medication for hypertension in primary care: a cluster randomized controlled trial [ISRCTN72772651]. *Implement Sci*. Sep 17, 2016;11(1):124. [doi: [10.1186/s13012-016-0486-3](#)] [Medline: [27640126](#)]

18. Singh S, Jamal A, Qureshi F. Readability metrics in patient education: where do we innovate? *Clin Pract*. Nov 4, 2024;14(6):2341-2349. [doi: [10.3390/clinpract14060183](https://doi.org/10.3390/clinpract14060183)] [Medline: [39585011](https://pubmed.ncbi.nlm.nih.gov/39585011/)]
19. Yu KH, Beam AL, Kohane IS. Artificial intelligence in healthcare. *Nat Biomed Eng*. Oct 2018;2(10):719-731. [doi: [10.1038/s41551-018-0305-z](https://doi.org/10.1038/s41551-018-0305-z)] [Medline: [31015651](https://pubmed.ncbi.nlm.nih.gov/31015651/)]
20. Alam SF, Gonzalez Suarez ML. Transforming healthcare: the AI revolution in the comprehensive care of hypertension. *Clin Pract*. Jul 10, 2024;14(4):1357-1374. [doi: [10.3390/clinpract14040109](https://doi.org/10.3390/clinpract14040109)] [Medline: [39051303](https://pubmed.ncbi.nlm.nih.gov/39051303/)]
21. Layton AT. A heart-to-heart with ChatGPT: AI applications in hypertension. *Am J Hypertens*. Aug 14, 2025;38(9):621-627. [doi: [10.1093/ajh/hpaf045](https://doi.org/10.1093/ajh/hpaf045)] [Medline: [40176295](https://pubmed.ncbi.nlm.nih.gov/40176295/)]
22. Aydin S, Karabacak M, Vlachos V, Margetis K. Large language models in patient education: a scoping review of applications in medicine. *Front Med (Lausanne)*. 2024;11:1477898. [doi: [10.3389/fmed.2024.1477898](https://doi.org/10.3389/fmed.2024.1477898)] [Medline: [39534227](https://pubmed.ncbi.nlm.nih.gov/39534227/)]
23. Serugunda HM, Jianquan O, Kasujja Namatovu H, et al. Using large language models for chronic disease management tasks: scoping review. *JMIR Med Inform*. Sep 29, 2025;13:e66905. [doi: [10.2196/66905](https://doi.org/10.2196/66905)] [Medline: [41021927](https://pubmed.ncbi.nlm.nih.gov/41021927/)]
24. Pariente B, Varennes O, Burgun A, Azizi M, Amar L, Tsopra R. Empowering patients and clinicians: LLMs in hypertension care, a scoping review. *Hypertension*. Jul 2026;83(7):e27004. [doi: [10.1161/HYPERTENSIONAHA.126.27004](https://doi.org/10.1161/HYPERTENSIONAHA.126.27004)] [Medline: [42021742](https://pubmed.ncbi.nlm.nih.gov/42021742/)]
25. Peters MDJ, Marnie C, Tricco AC, et al. Updated methodological guidance for the conduct of scoping reviews. *JB Evid Synth*. Oct 2020;18(10):2119-2126. [doi: [10.11124/JBIES-20-00167](https://doi.org/10.11124/JBIES-20-00167)] [Medline: [33038124](https://pubmed.ncbi.nlm.nih.gov/33038124/)]
26. Pollock D, Peters MDJ, Khalil H, et al. Recommendations for the extraction, analysis, and presentation of results in scoping reviews. *JB Evid Synth*. Mar 1, 2023;21(3):520-532. [doi: [10.11124/JBIES-22-00123](https://doi.org/10.11124/JBIES-22-00123)] [Medline: [36081365](https://pubmed.ncbi.nlm.nih.gov/36081365/)]
27. Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Med Res Methodol*. Nov 19, 2018;18(1):143. [doi: [10.1186/s12874-018-0611-x](https://doi.org/10.1186/s12874-018-0611-x)] [Medline: [30453902](https://pubmed.ncbi.nlm.nih.gov/30453902/)]
28. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med*. Oct 2, 2018;169(7):467-473. [doi: [10.7326/M18-0850](https://doi.org/10.7326/M18-0850)] [Medline: [30178033](https://pubmed.ncbi.nlm.nih.gov/30178033/)]
29. Rethlefsen ML, Kirtley S, Waffenschmidt S, et al. PRISMA-S: an extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews. *Syst Rev*. Jan 26, 2021;10(1):39. [doi: [10.1186/s13643-020-01542-z](https://doi.org/10.1186/s13643-020-01542-z)] [Medline: [33499930](https://pubmed.ncbi.nlm.nih.gov/33499930/)]
30. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. Mar 29, 2021;372:n71. [doi: [10.1136/bmj.n71](https://doi.org/10.1136/bmj.n71)] [Medline: [33782057](https://pubmed.ncbi.nlm.nih.gov/33782057/)]
31. Attribution 4.0 international (CC BY 4.0). Creative Commons. URL: <https://creativecommons.org/licenses/by/4.0/> [Accessed 2026-06-27]
32. Hong QN, Gonzalez-Reyes A, Pluye P. Improving the usefulness of a tool for appraising the quality of qualitative, quantitative and mixed methods studies, the Mixed Methods Appraisal Tool (MMAT). *J Eval Clin Pract*. Jun 2018;24(3):459-467. [doi: [10.1111/jep.12884](https://doi.org/10.1111/jep.12884)] [Medline: [29464873](https://pubmed.ncbi.nlm.nih.gov/29464873/)]
33. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *Int J Soc Res Methodol*. Feb 2005;8(1):19-32. [doi: [10.1080/1364557032000119616](https://doi.org/10.1080/1364557032000119616)]
34. Persell SD, Peprah YA, Lipiszko D, et al. Effect of home blood pressure monitoring via a smartphone hypertension coaching application or tracking application on adults with uncontrolled hypertension: a randomized clinical trial. *JAMA Netw Open*. Mar 2, 2020;3(3):e200255. [doi: [10.1001/jamanetworkopen.2020.0255](https://doi.org/10.1001/jamanetworkopen.2020.0255)] [Medline: [32119093](https://pubmed.ncbi.nlm.nih.gov/32119093/)]
35. Griffin AC, Xing Z, Mikles SP, et al. Information needs and perceptions of chatbots for hypertension medication self-management: a mixed methods study. *JAMIA Open*. Apr 2021;4(2):ooab021. [doi: [10.1093/jamiaopen/ooab021](https://doi.org/10.1093/jamiaopen/ooab021)] [Medline: [33898936](https://pubmed.ncbi.nlm.nih.gov/33898936/)]
36. Kario K, Nomura A, Harada N, et al. Efficacy of a digital therapeutics system in the management of essential hypertension: the HERB-DH1 pivotal trial. *Eur Heart J*. Oct 21, 2021;42(40):4111-4122. [doi: [10.1093/eurheartj/ehab559](https://doi.org/10.1093/eurheartj/ehab559)] [Medline: [34455443](https://pubmed.ncbi.nlm.nih.gov/34455443/)]
37. Gutierrez MM, Sakulbumrungsil R. Effectiveness of a pharmacist-led expert system intervention for medication adherence and blood pressure control of adults with hypertension: a randomized controlled trial. *Res Social Adm Pharm*. Jun 2023;19(6):931-943. [doi: [10.1016/j.sapharm.2023.03.004](https://doi.org/10.1016/j.sapharm.2023.03.004)] [Medline: [36941159](https://pubmed.ncbi.nlm.nih.gov/36941159/)]
38. Griffin AC, Khairat S, Bailey SC, Chung AE. A chatbot for hypertension self-management support: user-centered design, development, and usability testing. *JAMIA Open*. Oct 2023;6(3):ooad073. [doi: [10.1093/jamiaopen/ooad073](https://doi.org/10.1093/jamiaopen/ooad073)] [Medline: [37693367](https://pubmed.ncbi.nlm.nih.gov/37693367/)]
39. Sakane N, Suganuma A, Domichi M, et al. The effect of a mHealth app (KENPO-app) for specific health guidance on weight changes in adults with obesity and hypertension: pilot randomized controlled trial. *JMIR Mhealth Uhealth*. Apr 12, 2023;11:e43236. [doi: [10.2196/43236](https://doi.org/10.2196/43236)] [Medline: [37043287](https://pubmed.ncbi.nlm.nih.gov/37043287/)]

40. Yano Y, Nishiyama A, Suzuki Y, et al. Relevance of ChatGPT's responses to common hypertension-related patient inquiries. *Hypertension*. Jan 2024;81(1):e1-e4. [doi: [10.1161/HYPERTENSIONAHA.123.22084](https://doi.org/10.1161/HYPERTENSIONAHA.123.22084)] [Medline: [37916418](https://pubmed.ncbi.nlm.nih.gov/37916418/)]
41. O'Hagan E, McIntyre D, Laranjo L. The potential for a chat-based artificial intelligence model to facilitate educational messaging on hypertension. *Hypertension*. Aug 2023;80(8):e128-e130. [doi: [10.1161/HYPERTENSIONAHA.123.21395](https://doi.org/10.1161/HYPERTENSIONAHA.123.21395)] [Medline: [37325936](https://pubmed.ncbi.nlm.nih.gov/37325936/)]
42. Almagazzachi A, Mustafa A, Eighaei Sedeh A, et al. Generative artificial intelligence in patient education: ChatGPT takes on hypertension questions. *Cureus*. Feb 2024;16(2):e53441. [doi: [10.7759/cureus.53441](https://doi.org/10.7759/cureus.53441)] [Medline: [38435177](https://pubmed.ncbi.nlm.nih.gov/38435177/)]
43. Lee TJ, Campbell DJ, Patel S, et al. Unlocking health literacy: the ultimate guide to hypertension education from ChatGPT versus Google Gemini. *Cureus*. May 2024;16(5):e59898. [doi: [10.7759/cureus.59898](https://doi.org/10.7759/cureus.59898)] [Medline: [38721479](https://pubmed.ncbi.nlm.nih.gov/38721479/)]
44. Vinufrancis A, Al Hussein H, Patel HV, et al. Assessing the quality and reliability of AI-generated responses to common hypertension queries. *Cureus*. Aug 2024;16(8):e66041. [doi: [10.7759/cureus.66041](https://doi.org/10.7759/cureus.66041)] [Medline: [39224724](https://pubmed.ncbi.nlm.nih.gov/39224724/)]
45. Leitner J, Chiang PH, Agnihotri P, Dey S. The effect of an AI-based, autonomous, digital health intervention using precise lifestyle guidance on blood pressure in adults with hypertension: single-arm nonrandomized trial. *JMIR Cardio*. May 28, 2024;8:e51916. [doi: [10.2196/51916](https://doi.org/10.2196/51916)] [Medline: [38805253](https://pubmed.ncbi.nlm.nih.gov/38805253/)]
46. Niko MM, Karbasi Z, Kazemi M, Zahmatkeshan M. Comparing ChatGPT and Bing, in response to the Home Blood Pressure Monitoring (HBPM) knowledge checklist. *Hypertens Res*. May 2024;47(5):1401-1409. [doi: [10.1038/s41440-024-01624-8](https://doi.org/10.1038/s41440-024-01624-8)] [Medline: [38438722](https://pubmed.ncbi.nlm.nih.gov/38438722/)]
47. Sun T, Xu X, Ding Z, et al. Development of a health behavioral digital intervention for patients with hypertension based on an intelligent health promotion system and WeChat: randomized controlled trial. *JMIR Mhealth Uhealth*. Apr 5, 2024;12:e53006. [doi: [10.2196/53006](https://doi.org/10.2196/53006)] [Medline: [38578692](https://pubmed.ncbi.nlm.nih.gov/38578692/)]
48. Xu Y, Liu Q, Pang J, et al. Assessment of personalized exercise prescriptions issued by ChatGPT 4.0 and intelligent health promotion systems for patients with hypertension comorbidities based on the transtheoretical model: a comparative analysis. *J Multidiscip Healthc*. 2024;17:5063-5078. [doi: [10.2147/JMDH.S477452](https://doi.org/10.2147/JMDH.S477452)] [Medline: [39539514](https://pubmed.ncbi.nlm.nih.gov/39539514/)]
49. Aguzzi G, Magnini M, Farahmand A, Ferretti S, Pengo MF, Montagna S. RAG-enhanced open SLMs for hypertension management chatbots. *J Med Syst*. Nov 13, 2025;49(1):159. [doi: [10.1007/s10916-025-02297-7](https://doi.org/10.1007/s10916-025-02297-7)] [Medline: [41231304](https://pubmed.ncbi.nlm.nih.gov/41231304/)]
50. Antia SE, Ugwu CN, Ghodka V, et al. Healthy Heart Assistant, a WhatsApp-based generative pretrained transformer technology, for self-care in hypertensive patients. *Mayo Clin Proc Digit Health*. Sep 2025;3(3):100243. [doi: [10.1016/j.mcpdig.2025.100243](https://doi.org/10.1016/j.mcpdig.2025.100243)] [Medline: [40686624](https://pubmed.ncbi.nlm.nih.gov/40686624/)]
51. Jelic A, Sesto I, Rotkvic L, et al. Evaluating the clinical effectiveness and patient experience of a large language model-based digital tool for home-based blood pressure management: mixed methods study. *JMIR Mhealth Uhealth*. Nov 3, 2025;13:e68361. [doi: [10.2196/68361](https://doi.org/10.2196/68361)] [Medline: [41183160](https://pubmed.ncbi.nlm.nih.gov/41183160/)]
52. Montagna S, Ferretti S, Klopfenstein LC, et al. Privacy-preserving LLM-based chatbots for hypertensive patient self-management. *Smart Health*. Jun 2025;36:100552. [doi: [10.1016/j.smhl.2025.100552](https://doi.org/10.1016/j.smhl.2025.100552)]
53. Moolsart S, Kritpolviman KM. Self-health monitoring by smart devices and ontology technology for older adults with uncontrolled hypertension: quasi-experimental study. *JMIR Nurs*. Oct 14, 2025;8:e73386. [doi: [10.2196/73386](https://doi.org/10.2196/73386)] [Medline: [41086427](https://pubmed.ncbi.nlm.nih.gov/41086427/)]
54. Wang Y, Zhu T, Zhou T, et al. Hyper-DREAM, a multimodal digital transformation hypertension management platform integrating large language model and digital phenotyping: multicenter development and initial validation study. *J Med Syst*. Apr 2, 2025;49(1):42. [doi: [10.1007/s10916-025-02176-1](https://doi.org/10.1007/s10916-025-02176-1)] [Medline: [40172683](https://pubmed.ncbi.nlm.nih.gov/40172683/)]
55. Wang Y, Luan Y, Cheng S, et al. A multi-layer retrieval-augmented large language model framework for enhancing hypertension education. *Hypertens Res*. Apr 2026;49(4):1428-1440. [doi: [10.1038/s41440-025-02481-9](https://doi.org/10.1038/s41440-025-02481-9)] [Medline: [41501362](https://pubmed.ncbi.nlm.nih.gov/41501362/)]
56. Wang Y, Tan W, Cheng S, et al. Large language model agent for managing patients with suspected hypertension. *Hypertension*. Jan 2026;83(1):212-224. [doi: [10.1161/HYPERTENSIONAHA.125.25305](https://doi.org/10.1161/HYPERTENSIONAHA.125.25305)] [Medline: [41064862](https://pubmed.ncbi.nlm.nih.gov/41064862/)]
57. Yao Q, Qiu B, He L, et al. Effects of artificial intelligence recognition-based telerehabilitation on exercise capacity in patients with hypertension: randomized controlled trial. *J Med Internet Res*. Jan 13, 2026;28:e81400. [doi: [10.2196/81400](https://doi.org/10.2196/81400)] [Medline: [41529829](https://pubmed.ncbi.nlm.nih.gov/41529829/)]
58. Mathews SC, McShea MJ, Hanley CL, Ravitz A, Labrique AB, Cohen AB. Digital health: a path to validation. *NPJ Digit Med*. 2019;2(1):38. [doi: [10.1038/s41746-019-0111-3](https://doi.org/10.1038/s41746-019-0111-3)] [Medline: [31304384](https://pubmed.ncbi.nlm.nih.gov/31304384/)]
59. Kreuter MW, Wray RJ. Tailored and targeted health communication: strategies for enhancing information relevance. *Am J Health Behav*. 2003;27 Suppl 3:S227-S232. [doi: [10.5993/ajhb.27.1.s3.6](https://doi.org/10.5993/ajhb.27.1.s3.6)] [Medline: [14672383](https://pubmed.ncbi.nlm.nih.gov/14672383/)]
60. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. *Health Care (Don Mills)*. Mar 19, 2023;11(6):887. [doi: [10.3390/healthcare11060887](https://doi.org/10.3390/healthcare11060887)]

61. Omar M, Sorin V, Wieler LH, et al. Mapping the susceptibility of large language models to medical misinformation across clinical notes and social media: a cross-sectional benchmarking analysis. *Lancet Digit Health*. Jan 2026;8(1):100949. [doi: [10.1016/j.landig.2025.100949](https://doi.org/10.1016/j.landig.2025.100949)] [Medline: [41672646](https://pubmed.ncbi.nlm.nih.gov/41672646/)]
62. Du S, Zhou Y, Fu C, Wang Y, Du X, Xie R. Health literacy and health outcomes in hypertension: an integrative review. *Int J Nurs Sci*. Jul 10, 2018;5(3):301-309. [doi: [10.1016/j.ijnss.2018.06.001](https://doi.org/10.1016/j.ijnss.2018.06.001)] [Medline: [31406840](https://pubmed.ncbi.nlm.nih.gov/31406840/)]
63. Magi CE, El Aoufy K, Amato C, et al. The association between self-care and health literacy in patients with chronic diseases: a systematic review and meta-analysis. *J Clin Nurs*. Jul 2026;35(7):3011-3029. [doi: [10.1111/jocn.70291](https://doi.org/10.1111/jocn.70291)] [Medline: [41854032](https://pubmed.ncbi.nlm.nih.gov/41854032/)]
64. Murali L, Gopakumar G, Viswanathan DM, Raman R, Nedungadi P. Integrating LLMs and knowledge graphs for medical AI: advances, challenges, and future directions. *IEEE J Biomed Health Inform*. May 2026;30(5):3833-3848. [doi: [10.1109/JBHI.2025.3622058](https://doi.org/10.1109/JBHI.2025.3622058)] [Medline: [41100224](https://pubmed.ncbi.nlm.nih.gov/41100224/)]
65. Parameswaran V, Bernard J, Bernard A, et al. Evaluating large language models and retrieval-augmented generation enhancement for delivering guideline-adherent nutrition information for cardiovascular disease prevention: cross-sectional study. *J Med Internet Res*. Oct 7, 2025;27:e78625. [doi: [10.2196/78625](https://doi.org/10.2196/78625)] [Medline: [41057043](https://pubmed.ncbi.nlm.nih.gov/41057043/)]
66. Woo JJ, Yang AJ, Olsen RJ, et al. Custom large language models improve accuracy: comparing retrieval augmented generation and artificial intelligence agents to noncustom models for evidence-based medicine. *Arthroscopy*. Mar 2025;41(3):565-573. [doi: [10.1016/j.arthro.2024.10.042](https://doi.org/10.1016/j.arthro.2024.10.042)] [Medline: [39521391](https://pubmed.ncbi.nlm.nih.gov/39521391/)]
67. Li Z, Liu H, Tan W, et al. The effects of multitype prompt engineering for large language models in hypertension treatment decisions. *NPJ Digit Med*. Apr 15, 2026. [doi: [10.1038/s41746-026-02645-y](https://doi.org/10.1038/s41746-026-02645-y)] [Medline: [41986562](https://pubmed.ncbi.nlm.nih.gov/41986562/)]
68. Roberts SLE, Healey A, Sevdalis N. Use of health economic evaluation in the implementation and improvement science fields—a systematic literature review. *Implement Sci*. Jul 15, 2019;14(1):72. [doi: [10.1186/s13012-019-0901-7](https://doi.org/10.1186/s13012-019-0901-7)] [Medline: [31307489](https://pubmed.ncbi.nlm.nih.gov/31307489/)]
69. Malhotra A, Thompson RR, Kagoya F, et al. Economic evaluation of implementation science outcomes in low- and middle-income countries: a scoping review. *Implement Sci*. Nov 16, 2022;17(1):76. [doi: [10.1186/s13012-022-01248-x](https://doi.org/10.1186/s13012-022-01248-x)] [Medline: [36384807](https://pubmed.ncbi.nlm.nih.gov/36384807/)]
70. Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K. Low health literacy and health outcomes: an updated systematic review. *Ann Intern Med*. Jul 19, 2011;155(2):97-107. [doi: [10.7326/0003-4819-155-2-201107190-00005](https://doi.org/10.7326/0003-4819-155-2-201107190-00005)] [Medline: [21768583](https://pubmed.ncbi.nlm.nih.gov/21768583/)]
71. Craig P, Dieppe P, Macintyre S, et al. Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ*. Sep 29, 2008;337:a1655. [doi: [10.1136/bmj.a1655](https://doi.org/10.1136/bmj.a1655)] [Medline: [18824488](https://pubmed.ncbi.nlm.nih.gov/18824488/)]
72. Airhihenbuwa CO, Ford CL, Iwelunmor JI. Why culture matters in health interventions: lessons from HIV/AIDS stigma and NCDs. *Health Educ Behav*. Feb 2014;41(1):78-84. [doi: [10.1177/1090198113487199](https://doi.org/10.1177/1090198113487199)] [Medline: [23685666](https://pubmed.ncbi.nlm.nih.gov/23685666/)]
73. Miezah D, Hayman LL. Culturally tailored lifestyle modification strategies for hypertension management: a narrative review. *Am J Lifestyle Med*. Jan 2026;20(1):46-54. [doi: [10.1177/15598276241297675](https://doi.org/10.1177/15598276241297675)] [Medline: [39540161](https://pubmed.ncbi.nlm.nih.gov/39540161/)]
74. Yardley L, Spring BJ, Riper H, et al. Understanding and promoting effective engagement with digital behavior change interventions. *Am J Prev Med*. Nov 2016;51(5):833-842. [doi: [10.1016/j.amepre.2016.06.015](https://doi.org/10.1016/j.amepre.2016.06.015)] [Medline: [27745683](https://pubmed.ncbi.nlm.nih.gov/27745683/)]
75. Müller A, Schaaf J. User or patient/human or person? Development of a practical framework for applying user-centered design to vulnerable populations for digital transformation in healthcare. *Digit Health*. 2025;11:20552076251375835. [doi: [10.1177/20552076251375835](https://doi.org/10.1177/20552076251375835)] [Medline: [40933080](https://pubmed.ncbi.nlm.nih.gov/40933080/)]
76. Tucker KL, Sheppard JP, Stevens R, et al. Self-monitoring of blood pressure in hypertension: a systematic review and individual patient data meta-analysis. *PLoS Med*. Sep 2017;14(9):e1002389. [doi: [10.1371/journal.pmed.1002389](https://doi.org/10.1371/journal.pmed.1002389)] [Medline: [28926573](https://pubmed.ncbi.nlm.nih.gov/28926573/)]

Abbreviations

AI: artificial intelligence

LLM: large language model

ML: machine learning

MMAT: Mixed Methods Appraisal Tool

NLP: natural language processing

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews

RAG: retrieval-augmented generation

RCT: randomized controlled trial

Edited by Stefano Brini; peer-reviewed by Xin Li, Yijun Wang; submitted 18.Mar.2026; final revised version received 14.Jun.2026; accepted 15.Jun.2026; published 15.Jul.2026

Please cite as:

Chen H, Xiao S, Wan T, Li G, Peng Y, Wang Z

Application of AI in Hypertension Health Education: Scoping Review

J Med Internet Res 2026;28:e95596

URL: <https://www.jmir.org/2026/1/e95596>

doi: [10.2196/95596](https://doi.org/10.2196/95596)

© Haoran Chen, Shenglan Xiao, Tong Wan, Gui Li, Yanhong Peng, Zhimin Wang. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 15.Jul.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 the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.