ARTIFICIAL INTELLIGENCE (AI) AND DIGITAL HEALTHCARE

RecoverAI: A Context-Grounded WhatsApp Conversational Assistant for Personalized Post-Discharge Patient Follow-Up

Praharsha Ravindra Ekshinge, Shukranti Dilip Mane, Drushti Dadaso Desai, Aishwarya Rajiv Banage, Supriya S. Dongre

Department of Computer Science & Engineering, Sanjay Ghodawat Institute, Kolhapur, Maharashtra, India

Published November 2026

Abstract

Post-discharge care requires patients to follow medication instructions, attend appointments, understand recovery plans, and communicate concerns after leaving a hospital. This literature review examines digital follow-up, mobile medication-adherence interventions, healthcare conversational agents, and large language model (LLM) safety to identify design requirements for RecoverAI. The reviewed studies show that digital communication can support continuity and patient engagement, but effects on readmissions and adherence vary by intervention and population. Research on conversational agents also highlights the need to evaluate usability, information quality, safety, and clinical outcomes rather than treating message exchange as evidence of effectiveness. RecoverAI is an academic prototype that connects a React and Vite staff dashboard, Supabase PostgreSQL records, a Supabase Edge Function, an OpenWA WhatsApp gateway, and Gemini-assisted response generation. Demonstration records confirm retrieval of medication name, dosage, frequency, and instructions from a registered patient record, and delivery of an appointment reminder to a test phone. These observations establish selected integration paths only; they do not establish clinical benefit, complete medication-reminder scheduling, multilingual validation, or production readiness. The review identifies research needs in patient identity verification, grounded response accuracy, reliable message processing, accessible language support, privacy, and outcome-based evaluation.

Keywords: post-discharge care; digital health; conversational AI; WhatsApp; medication adherence; patient follow-up

1 Introduction

Hospital discharge is a transition from supervised care to self-management. Patients may need to take medicines at specified times, attend follow-up appointments, monitor symptoms, and follow recovery instructions. Questions can arise after the patient returns home, while hospital staff must coordinate follow-up across many patients. Paper instructions, telephone calls, SMS messages, and separate patient portals each address part of this process, but they may create fragmented communication and additional work for staff.

Digital health interventions offer ways to support communication after discharge. However, the existence of a digital channel does not itself demonstrate better health outcomes. Reviews of health-information-technology discharge interventions report varied results, and randomized studies of automated messages and video follow-up have not consistently shown reductions in acute-care use or improvements in all adherence measures [1, 2]. Therefore, a follow-up system should be evaluated

for what it actually does: correct patient identification, accurate retrieval of recorded information, timely delivery, safe handling of questions, and usefulness to patients and staff.

RecoverAI explores an integrated workflow in which hospital staff maintain patient information through an existing web dashboard and registered patients communicate through WhatsApp. The backend retrieves relevant structured records before the generative AI service is asked to phrase a response. The database, rather than the language model, remains the intended source of patient-specific facts. The project is positioned as a workflow-support prototype, not a diagnostic tool or a replacement for clinical judgement.

The objectives of this literature review are to: (1) examine research on digital post-discharge follow-up and medication adherence; (2) review healthcare conversational agents and LLM evaluation; (3) compare selected studies by method, contribution, and limitation; and (4) identify research gaps relevant to RecoverAI's design and future evaluation.

2 Literature review

Review approach: This paper uses a focused narrative review rather than a systematic review. The selected literature includes peer-reviewed systematic reviews, scoping reviews, meta-analyses, randomized trials, and a digital-health guideline. Studies were selected for their relevance to four themes: post-discharge continuity, mobile medication adherence, healthcare conversational agents, and LLM evaluation and safety. Most selected research was published from 2019 to 2026; a small number of earlier studies are retained as foundational context. The comparison is qualitative because the reviewed studies use different populations, interventions, outcomes, and follow-up periods. No pooled effect estimate is calculated in this paper.

2.1 Digital follow-up and continuity of care

Abraham et al. reviewed health-information-technology interventions used during discharge transitions and assessed their association with readmissions and emergency-department visits [1]. Their work considered technology-supported care-transition interventions rather than a single messaging product. It is relevant because RecoverAI also links communication with a patient's existing records. The review's broad intervention scope means that its findings cannot be transferred directly to a WhatsApp chatbot. The review is useful for framing the problem as a care-transition workflow rather than a messaging problem alone. Its synthesis indicates why future evaluation should distinguish process measures, such as whether a patient receives and answers a message, from clinical measures, such as readmission or emergency-department use. The interventions reviewed were not identical, so a single effect estimate cannot be assumed for RecoverAI.

Bressman et al. conducted a randomized clinical trial of an automated text-message programme for 30 days after hospital discharge [2]. Patients received check-in messages, and identified needs could be escalated to practice staff. The trial found no significant reduction in 30-day acute-care revisits compared with usual transitional-care calls, although many participants responded to at least one message. This study demonstrates that engagement and clinical outcomes are separate measures. The programme used repeated check-ins and routed identified needs to practice staff. Although most intervention participants answered at least one message, the trial did not show a significant reduction in acute-care revisits within 30 days. This is an important comparison for RecoverAI because it demonstrates that a functioning messaging workflow and user engagement do not necessarily translate into reduced hospital utilization.

A randomized trial of post-discharge video visits examined whether provider-led video review improved adherence to hospital discharge recommendations [3]. The intervention did not significantly improve the overall primary adherence outcome compared with usual care. The result cautions against assuming that adding a digital contact automatically improves adherence; the content, timing, patient population, and care context matter. The study included medication management, self-management and home-support recommendations in its adherence outcome. The absence of a significant overall difference suggests that adding a follow-up modality should be assessed against usual care and should not be described as effective solely because it is technically available.

Sekhon et al. reviewed digital technologies used for post-discharge follow-up and management in older adults, including video visits, mobile applications, and remote monitoring [4]. The review reported feasibility and generally positive engagement and satisfaction, while findings for readmission and mental-health outcomes were mixed. Its population focus is older adults, so its conclusions should not be generalized to every age group or messaging platform. The review considered feasibility, acceptability, adherence, satisfaction, and health-related outcomes. Its mixed findings for rehospitalization and mental-health outcomes reinforce the need for population-specific evaluation. RecoverAI should therefore report its intended patient population and avoid generalizing results from older-adult studies to all patients.

Woods et al. synthesized evidence on nurse-led post-discharge telephone follow-up [5]. Their review described variation in intervention design and outcome measurement. Mistiaen and Poot's earlier Cochrane review also reported heterogeneous evidence for telephone follow-up after discharge [6]. These studies provide historical context for digital follow-up: the communication channel is only one part of a wider care process.

2.2 Mobile health and medication adherence

Armitage et al. reviewed randomized trials of mobile applications designed to support medication adherence [7]. The review examined whether app-based strategies improved adherence and emphasized the need to interpret results in relation to intervention design and trial quality. Kim et al. later conducted a systematic review and meta-analysis of mobile-health interventions for medication adherence [8]. Their findings support the potential of mobile health, but do not imply that every reminder schedule or implementation will be effective. Their review provides evidence relevant to digital adherence support, but mobile applications differ in reminders, monitoring, feedback, and interaction design. RecoverAI currently retrieves recorded medicine details; the demonstration does not show that patients took medicines as prescribed or that reminders improved adherence.

The literature supports investigating reminders as one component of follow-up, but adherence depends on more than message delivery. Patients may have changing prescriptions, multiple medicines, different routines, limited connectivity, or difficulty understanding instructions. RecoverAI's demonstration confirmed retrieval of medication name, dosage, frequency, and instructions from a patient record. It did not establish adherence improvement or comprehensive reminder performance across schedules, retries, and time zones.

2.3 Healthcare conversational agents

Laranjo et al. systematically reviewed conversational agents in healthcare and described applications using natural-language input [9]. They noted limitations in the evidence base and the need for stronger reporting and evaluation. Car et al. conducted a scoping review and conceptual analysis of

conversational agents in healthcare [10], highlighting the diversity of use cases and the need to assess acceptability, safety, and effectiveness. These reviews show that a chatbot interface alone is not a sufficient research contribution. The review also highlighted limitations in the evidence base and the need for stronger evaluation. For RecoverAI, this supports specifying the exact task of the agent, answering questions about recorded recovery information rather than presenting a general-purpose medical chatbot.

Milne-Ives et al. reviewed the effectiveness of AI conversational agents in healthcare [11]. Reported outcomes were positive or mixed, while limitations in study quality constrained the strength of conclusions. Li et al. reviewed randomized controlled trials of AI-driven conversational agents [12]. They reported that agents were feasible and acceptable in several contexts, but effects varied across outcomes. The results support cautious exploration of conversational support as an adjunct to existing services.

Denecke and May reviewed usability assessment of healthcare conversational agents and observed that usability evaluation lacks a single standard procedure [13]. Denecke proposed a framework for guiding development and evaluation of high-quality healthcare conversational agents [14]. Ding et al. reviewed evaluation frameworks and identified stages covering feasibility/usability, efficacy, effectiveness, and implementation [15]. Together, these studies suggest that RecoverAI should be evaluated beyond successful message exchange, including information quality, user experience, safety, and workflow fit.

2.4 Large language models and healthcare safety

Wang et al. reviewed applications and concerns involving ChatGPT and other conversational LLMs in healthcare [16]. Their review grouped applications such as summarization and medical knowledge inquiry, while identifying reliability, bias, privacy, and public acceptability as concerns. Bedi et al. systematically reviewed evaluation of healthcare applications of LLMs [17]. They found substantial variation in evaluation methods and limited use of real patient-care data in the reviewed studies. Their synthesis is relevant to the separation between information retrieval and natural-language generation. Patient-specific medication details should come from an authorized record, while the model should be constrained to phrasing that information. This design reduces dependence on model memory but cannot, by itself, guarantee that the generated wording is faithful.

Huo et al. reviewed studies evaluating LLM chatbots for health advice [18]. The review identified inconsistent reporting of model characteristics and evaluation methods, as well as limited reporting of ethical, regulatory, and patient-safety implications. These findings are directly relevant to a patient-facing assistant. A fluent answer must not be treated as proof of correctness. RecoverAI therefore uses retrieved patient records as the intended source of individual facts, while still requiring tests for unsupported answers, missing data, unsafe requests, and incorrect patient matching.

The World Health Organization's guidance on digital interventions emphasizes that digital tools should complement a functioning health system and be considered in relation to benefits, harms, acceptability, feasibility, resources, and equity [19]. For RecoverAI, this means that patient consent, opt-out options, accessible language, secure information handling, and clinical escalation should be treated as core design requirements.

2.5 Synthesis of the reviewed literature

Across the discharge-transition studies, communication tools can extend contact beyond the hospital, but the measured effects are not uniform. Automated SMS can achieve patient engagement without necessarily reducing acute-care revisits [2], while video follow-up did not significantly improve the overall adherence outcome in the cited trial [3]. The digital-health review in older adults also reports mixed results for some clinical outcomes [4]. These findings support an evaluation strategy that separates technical reach, patient response, care-process completion, and clinical outcomes.

A second theme is that conversational quality and clinical safety must be evaluated together. Usability and evaluation-framework studies recommend staged evaluation rather than relying on a single demonstration [13–15]. LLM reviews further identify weaknesses in reporting, model transparency, and safety assessment [16–18]. For RecoverAI, the most defensible design choice is therefore to retrieve patient-specific facts from the authorized database and constrain the model to explain those facts. This design is a risk-reduction approach, not a guarantee against hallucination, incorrect matching, or unsafe interpretation.

3 Comparison of reviewed research

Table 1 summarizes 18 research papers selected for this narrative review. It records each study's method or technology, main contribution, and a limitation relevant to interpreting its findings. The table is an analytical synthesis; it does not imply that all studies evaluated the same population or intervention.

Table 1. Comparison of selected research papers.

Author(s), year	Method / technology; contribution; limitation
Abraham et al., 2022 [1]	Systematic review of HIT-based discharge-transition interventions. Synthesizes technology-supported care-transition approaches. Limitation: heterogeneous interventions; no WhatsApp-specific effectiveness evidence.
Bressman et al., 2024 [2]	Randomized trial of automated post-discharge SMS check-ins with staff escalation. Tests scalable follow-up. Limitation: no significant reduction in 30-day acute-care revisits.
Dugani et al., 2023 [3]	Randomized trial of provider-led video visits reviewing discharge recommendations. Evaluates adherence after discharge. Limitation: no significant overall improvement in the primary adherence outcome.
Sekhon et al., 2026 [4]	Systematic review of video, mHealth and remote monitoring for older adults. Synthesizes feasibility and outcomes. Limitation: older-adult focus; mixed readmission and mental-health findings.
Woods et al., 2019 [5]	Mixed-studies systematic review of nurse-led post-discharge calls. Describes continuity-of-care models. Limitation: varied call methods and outcome measures.
Mistiaen & Poot, 2006 [6]	Cochrane review of professional telephone follow-up, post-discharge evidence. Provides foundational post-discharge evidence. Limitation: older, heterogeneous evidence and inconclusive overall effects.
Armitage et al., 2020 [7]	Systematic review/meta-analysis of mobile adherence apps. Assesses app-based medication support. Limitation: app interventions differ and do not validate a specific reminder workflow.
Kim et al., 2025 [8]	Systematic review/meta-analysis of mHealth adherence interventions. Synthesizes mobile-health evidence. Limitation: intervention and population variation limits direct transfer.
Laranjo et al., 2018 [9]	Systematic review of healthcare conversational agents. Maps natural-language applications. Limitation: evidence quality and reporting limitations; earlier than preferred window.

Author(s), year	Method / technology; contribution; limitation
Car et al., 2020 [10]	Scoping review and conceptual analysis of healthcare agents. Clarifies application areas. Limitation: broad scope; does not validate patient-record integration.
Milne-Ives et al., 2020 [11]	Systematic review of AI conversational agents. Reviews effectiveness across healthcare uses. Limitation: study quality and outcome heterogeneity limit firm conclusions.
Li et al., 2023 [12]	Systematic review of randomized trials of AI agents. Examines feasibility and effects. Limitation: varied outcomes; not limited to post-discharge care.
Denecke & May, 2022 [13]	Literature review of healthcare-agent usability. Highlights usability as a design requirement. Limitation: no single standardized usability procedure.
Denecke, 2023 [14]	Framework for developing and evaluating healthcare agents. Provides quality categories. Limitation: framework guidance alone does not prove clinical effectiveness.
Ding et al., 2024 [15]	Systematic scoping review of conversational-agent evaluation. Proposes staged evaluation. Limitation: framework needs validation in specific settings.
Wang et al., 2024 [16]	Systematic review of conversational LLM uses and concerns. Summarizes reliability, bias and privacy risks. Limitation: evidence for complex clinical tasks remains limited.
Bedi et al., 2025 [17]	Systematic review of healthcare LLM testing. Identifies variation in evaluation and limited real-care data. Limitation: many studies use non-clinical settings.
Huo et al., 2025 [18]	Systematic review of LLM chatbot health-advice studies. Examines reporting and safety. Limitation: model details and evaluation standards are often insufficient.

4 Research gap

The literature shows substantial work on digital discharge support, medication adherence, conversational agents, and LLM evaluation. However, the reviewed studies do not remove the need to investigate how these elements operate together in a hospital workflow. The following gaps are derived from the reviewed evidence and the current maturity of the RecoverAI prototype.

4.1 Integrated workflow evaluation

Many studies examine one intervention type, such as SMS, video calls, apps, or conversational agents. A hospital workflow must connect staff-entered records, patient identity, context retrieval, message generation, delivery, and staff visibility. RecoverAI implements a prototype of this chain, but its effect on staff workload and continuity of care has not been measured.

4.2 Patient identity and grounded responses

A record-grounded answer depends on identifying the correct patient and retrieving current, authorized information. Phone-number normalization and identifier resolution are practical integration steps, but false matching could disclose another patient's information. Evaluation should measure match precision, retrieval correctness, answer faithfulness, and refusal when context is missing.

4.3 Reliability across message processing

Webhook acceptance is not equivalent to successful delivery of one correct message. Duplicate events, retries, session expiry, network interruption, and AI-service failures require separate testing. Duplicate database insert handling is present in the prototype, but end-to-end exactly-once delivery has not been established.

4.4 Multilingual accessibility

Faculty feedback identified Marathi and Hindi as desired languages. The available project record does not confirm complete multilingual support. Language detection or preference selection, medical terminology, literacy, code-switching, and dosage-translation accuracy require validation with native speakers and healthcare professionals.

4.5 Privacy, safety, and clinical governance

The project has no documented security audit, clinical validation, regulatory assessment, or production privacy review. Consent, opt-out, identity verification, least-privilege staff access, safe message wording, retention rules, and clinician-approved escalation are prerequisites for real-world use.

4.6 Outcome-based evaluation

The demonstration does not measure medication adherence, missed appointments, readmissions, patient comprehension, staff time, satisfaction, or cost-effectiveness. These outcomes require a defined study design and suitable comparison group.

5 Proposed work: RecoverAI system

RecoverAI is designed as a two-interface post-discharge communication prototype. Hospital staff use the existing React and Vite dashboard to manage patient records, recovery plans, medication entries, and appointments. Patients use WhatsApp to receive supported reminders and ask questions about information recorded for them. Supabase PostgreSQL stores the structured data and message history. A Supabase Edge Function processes incoming webhook events, normalizes sender identifiers, matches a registered patient, retrieves relevant context, calls the Gemini API for natural-language phrasing, and sends the reply through the OpenWA gateway.

The database entities documented for the project are patients, recovery_plans, medications, appointments, and whatsapp_messages. The database is intended to remain the source of truth for patient-specific facts. The AI component should phrase retrieved information; it should not independently prescribe treatment, diagnose a condition, or modify a medication instruction. If the relevant information is missing or the AI service is unavailable, the assistant should communicate that it cannot verify the information and direct urgent concerns to the healthcare team or emergency services.

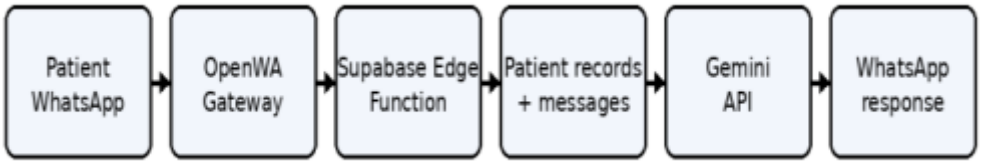


Figure 1. RecoverAI message-processing and context-retrieval architecture.

Figure 1 shows the message-processing and context-retrieval architecture. The incoming-message sequence is: (1) patient sends a WhatsApp message; (2) OpenWA receives it; (3) the gateway forwards a webhook event; (4) the Edge Function parses and normalizes the sender; (5) the backend resolves supported identifier variations; (6) the patient record is matched; (7) relevant patient,

recovery-plan, or medication context is retrieved; (8) Gemini generates a response using that context; and (9) the reply and message history are stored. Duplicate insert conflicts with code 23505 were detected and ignored in recorded development logs. This is a database-level safeguard, not proof of complete end-to-end idempotency.

The documented demonstration confirmed retrieval of medication name, dosage, frequency, and instructions from a registered patient record and successful delivery of an appointment reminder to a test phone. Some webhook requests returned HTTP 200 with EVENT_RECEIVED. Development records also mention multiple replies, patient-matching failures, recovery-plan lookup failures, and AI-service errors; the available evidence does not establish that every issue was fully eliminated. Complete medication-reminder scheduling, retry behaviour, time-zone handling, multilingual support, and clinical validation remain verification or future-work items.

6 Conclusion

This literature review examined digital post-discharge follow-up, mobile medication-adherence interventions, healthcare conversational agents, and LLM evaluation. Prior research indicates that digital communication can support contact and engagement, but outcomes vary and cannot be inferred from message delivery alone. The reviewed literature emphasizes usability, information quality, safety, privacy, and staged evaluation. RecoverAI applies these principles through a prototype that connects a staff dashboard, structured patient records, a WhatsApp gateway, and context-assisted response generation. Demonstrations confirmed selected medication-record retrieval and appointment-reminder delivery. These results establish technical feasibility for specific paths only; they do not establish clinical effectiveness, improved adherence, reduced readmissions, or production reliability.

7 Future scope

Future work should first strengthen patient identity verification and authorization, then test record retrieval and answer faithfulness using controlled test cases. Message processing should be evaluated for duplicate events, retries, ordering, delivery receipts, gateway disconnection, and recovery. Reminder scheduling should be verified across medication schedules and time zones. Marathi and Hindi support should be developed and tested with native speakers and healthcare professionals. The existing dashboard can be extended to display patient alerts and message-related follow-up without redesigning its current interface. Before any patient deployment, the project requires a privacy and security review, patient consent and opt-out procedures, clinician-approved escalation rules, a stable and policy-compliant messaging gateway, and appropriate institutional approvals. A future pilot should measure usability, comprehension, staff time, reminder timeliness, adherence, and patient outcomes using a defined comparator. A staged evaluation is recommended: begin with offline test cases and staff review, proceed to supervised usability testing, and only then consider a prospective pilot after institutional approvals. The first stage should prioritize false patient matches and unsupported medication answers because these can create direct privacy or safety risks. The later pilot should predefine primary outcomes and report participant characteristics, follow-up duration, missing data, and adverse events.

The proposed evaluation should report measurable indicators: patient-match precision and false-match rate; record-retrieval correctness; answer faithfulness and unsupported-answer rate; duplicate-event rate and delivery success; reminder timeliness; comprehension in English, Marathi

and Hindi; task-completion time; and patient/staff satisfaction. Clinical outcomes such as adherence, missed appointments and readmissions should be studied only in a prospective design with an appropriate comparator and required approvals.

References

[1] Abraham J, et al. Effect of health information technology (HIT)-based discharge transition interventions on patient readmissions and emergency room visits: A systematic review. *J Amer Med Inform Assoc*. 2022;29(4):735–748. https://doi.org/10.1093/jamia/ocac013

[2] Bressman E, et al. Automated text message-based program and use of acute health care resources after hospital discharge: A randomized clinical trial. *JAMA Netw Open*. 2024;7(4):e243701. https://doi.org/10.1001/jamanetworkopen.2024.3701

[3] Dugani SB, et al. Postdischarge video visits for adherence to hospital discharge recommendations: A randomized clinical trial. *Mayo Clin Proc Digit Health*. 2023;1(3):368–378. https://doi.org/10.1016/j.mcpdig.2023.06.006

[4] Sekhon H, et al. Effectiveness of digital health technologies for post-discharge follow-up and management in older adults: A systematic review. *Lancet Healthy Longev*. 2026;7(8):100885. https://doi.org/10.1016/j.lanhl.2026.100885

[5] Woods CE, Jones R, O'Shea E, Grist E, Wiggers J, Usher K. Nurse-led postdischarge telephone follow-up calls: A mixed study systematic review. *Journal of Clinical Nursing*. 2019;28(19–20):3386–3399. https://doi.org/10.1111/jocn.14951

[6] Mistiaen P, Poot E. Telephone follow-up, initiated by a hospital-based health professional, for postdischarge problems in patients discharged from hospital to home. *Cochrane Database of Systematic Reviews*. 2006;(4):CD004510. https://doi.org/10.1002/14651858.CD004510.pub3

[7] Armitage LC, Kassavou A, Sutton S. Do mobile device apps designed to support medication adherence demonstrate efficacy? A systematic review of randomised controlled trials, with meta-analysis. *BMJ Open*. 2020;10(1):e032045. https://doi.org/10.1136/bmjopen-2019-032045

[8] Kim SK, Park SY, Hwang HR, Moon SH, Park JW. Effectiveness of mobile health intervention in medication adherence: A systematic review and meta-analysis. *Journal of Medical Systems*. 2025;49(1):13. https://doi.org/10.1007/s10916-024-02135-2

[9] Laranjo L, et al. Conversational agents in healthcare: A systematic review. *J Amer Med Inform Assoc*. 2018;25(9):1248–1258. https://doi.org/10.1093/jamia/ocy072

[10] Tudor Car L, Dhinagaran DA, Kyaw BM, Kowatsch T, Joty S, Theng YL, et al. Conversational agents in health care: Scoping review and conceptual analysis. *Journal of Medical Internet Research*. 2020;22(8):e17158. https://doi.org/10.2196/17158

[11] Milne-Ives M, et al. The effectiveness of artificial intelligence conversational agents in health care: Systematic review. *J Med Internet Res*. 2020;22(10):e20346. https://doi.org/10.2196/20346

[12] Li Y, et al. Feasibility and effectiveness of artificial intelligence-driven conversational agents in healthcare interventions: A systematic review of randomized controlled trials. *Int J Nurs Stud*. 2023;143:104494. https://doi.org/10.1016/j.ijnurstu.2023.104494

[13] Denecke K, May R. Usability assessment of conversational agents in healthcare: A literature review. *Stud Health Technol Inform*. 2022;294:169–173. https://doi.org/10.3233/SHTI220431

[14] Denecke K. Framework for guiding the development of high-quality conversational agents in healthcare. *Healthcare*. 2023;11(8):1061. https://doi.org/10.3390/healthcare11081061

[15] Ding H, Simmich J, Vaezipour A, Andrews N, Russell T. Evaluation framework for conversational agents with artificial intelligence in health interventions: A systematic scoping review. *J Amer Med Inform Assoc*. 2024;31(3):746–761. https://doi.org/10.1093/jamia/ocad222

[16] Wang L, et al. Applications and concerns of ChatGPT and other conversational large language models in health care: Systematic review. *J Med Internet Res*. 2024;26:e22769. https://doi.org/10.2196/22769

[17] Bedi S, et al. Testing and evaluation of health care applications of large language models: A systematic review. *JAMA*. 2025;333(4):319–328. https://doi.org/10.1001/jama.2024.21700

[18] Huo B, et al. Large language models for chatbot health advice studies: A systematic review. *JAMA Netw Open*. 2025;8(2):e2457879. https://doi.org/10.1001/jamanetworkopen.2024.57879

[19] World Health Organization. *WHO guideline: Recommendations on digital interventions for health system strengthening*. Geneva: WHO; 2019. ISBN 978-92-4-155050-5. Available from: https://www.who.int/publications/i/item/9789241550505