

REVIEW

Open access

Artificial Intelligence for Digital Patient Communication: A Systematic Review of Portal Message Triage, Chatbot Support, Care Navigation, Automated Response Drafting, and Patient Engagement Analytics

Rania Hassan^{1*}, Dina Fathy¹

Abstract

Asynchronous digital communication with patients has become a routine component of modern healthcare delivery. The rapid growth of patient portals, chatbots, and digital front-door tools has created opportunities for more responsive care, while also increasing communication workload for clinical teams. This systematic review examined artificial intelligence applications in digital patient communication from 2017 to 2026. The review focused on portal message triage, chatbot support, care navigation, automated response drafting, and patient engagement analytics. A PRISMA 2020-compliant review was conducted using structured searches of PubMed, Scopus, IEEE Xplore, and Web of Science. Records were screened by two reviewers, with data extracted on communication domain, AI approach, clinical setting, evaluation strategy, safety reporting, and implementation maturity. The literature was dominated by studies of chatbot support and portal message triage, with a growing body of work on large language model-enabled response drafting. Care navigation and patient engagement analytics were less frequently evaluated, and most studies emphasized technical performance, user satisfaction, or feasibility rather than health outcomes or workload reduction in real-world settings. AI for patient communication appears technically promising in isolated tasks, particularly message classification, chatbot interaction, and draft response generation. However, evidence remains limited regarding safe, equitable, and effective deployment across integrated communication workflows.

Keywords Artificial intelligence, Large language models, Patient portals, Care navigation, Secure messaging, Chatbots

*Correspondence:

Rania Hassan
rania.hassan@gmail.com

¹ Department of Health Informatics and AI Systems, Faculty of Medicine, Alexandria University, Alexandria, Egypt

Introduction

Digital patient communication has become a defining feature of contemporary healthcare, driven by the expansion of patient portals, secure messaging, virtual care, and consumer-facing conversational tools. Early work on patient portal messages showed that digital communication contains clinically meaningful information and can be analyzed using computational methods to identify message type, intent, and resolution pathways [1-3]. In parallel, conversational agents became increasingly visible as patient-facing tools for symptom checking, education, administrative support, and

behavioral health support [4, 5]. Together, these developments positioned digital communication as both a care-delivery channel and a data source for artificial intelligence-enabled support.

The growth of asynchronous messaging has created substantial operational pressure for clinicians and health systems, particularly as secure messages increasingly require triage, documentation, and response outside traditional visit structures. Studies of provider-to-patient messaging and portal billing policies reported rising message volumes, uneven distribution of communication work, and concerns about after-hours burden [6, 7]. Artificial intelligence has therefore been proposed as a way to classify messages, route requests, identify urgency, draft replies, and reduce repetitive communication labor without fully removing human oversight [8, 9]. This promise is especially salient in settings where inbox burden, delayed response, and fragmented routing can affect patient experience and clinician workload.

Despite rapid growth, AI tools for digital patient communication remain fragmented across distinct technical and clinical domains. Portal-message classifiers, chatbot systems, care navigation tools, draft-reply models, and engagement prediction algorithms are often evaluated separately, even though patients experience these functions as a connected communication pathway [10–13]. Recent studies of large language model-generated replies have further expanded the field by shifting attention from classification and automation toward communication quality, clinician review, personalization, and ethical preferences [14–17]. A unified synthesis is therefore needed to examine how these technologies collectively support, augment, or reshape patient-facing communication.

This review systematically synthesizes peer-reviewed evidence on artificial intelligence for digital patient communication. It follows PRISMA 2020 principles and focuses on five domains: portal message triage, chatbot support, care navigation, automated response drafting, and patient engagement analytics. The review also examines evaluation outcomes, safety oversight, health equity, and implementation maturity across included studies. By integrating work across these domains, the review aims to clarify the state of evidence and identify priorities for safer, more equitable, and more clinically useful AI-supported communication.

Materials and Methods

Search strategy

A structured search strategy was developed to identify peer-reviewed studies from 2017 through 2026 that examined artificial intelligence in patient-facing digital communication. Searches were conducted in PubMed, Scopus, IEEE Xplore, and Web of Science using combinations of terms related to artificial intelligence, machine learning, natural language processing, patient portals, secure messaging, chatbots, conversational agents, digital front doors, care navigation, automated response drafting, and patient engagement analytics. Search strings were informed by prior research on message classification, conversational agents, portal-message analytics, and AI-generated patient replies. The search also included domain-specific terms for urgency detection, message routing, symptom checking, scheduling, referral navigation, follow-up reminders, satisfaction prediction, adherence, and disengagement.

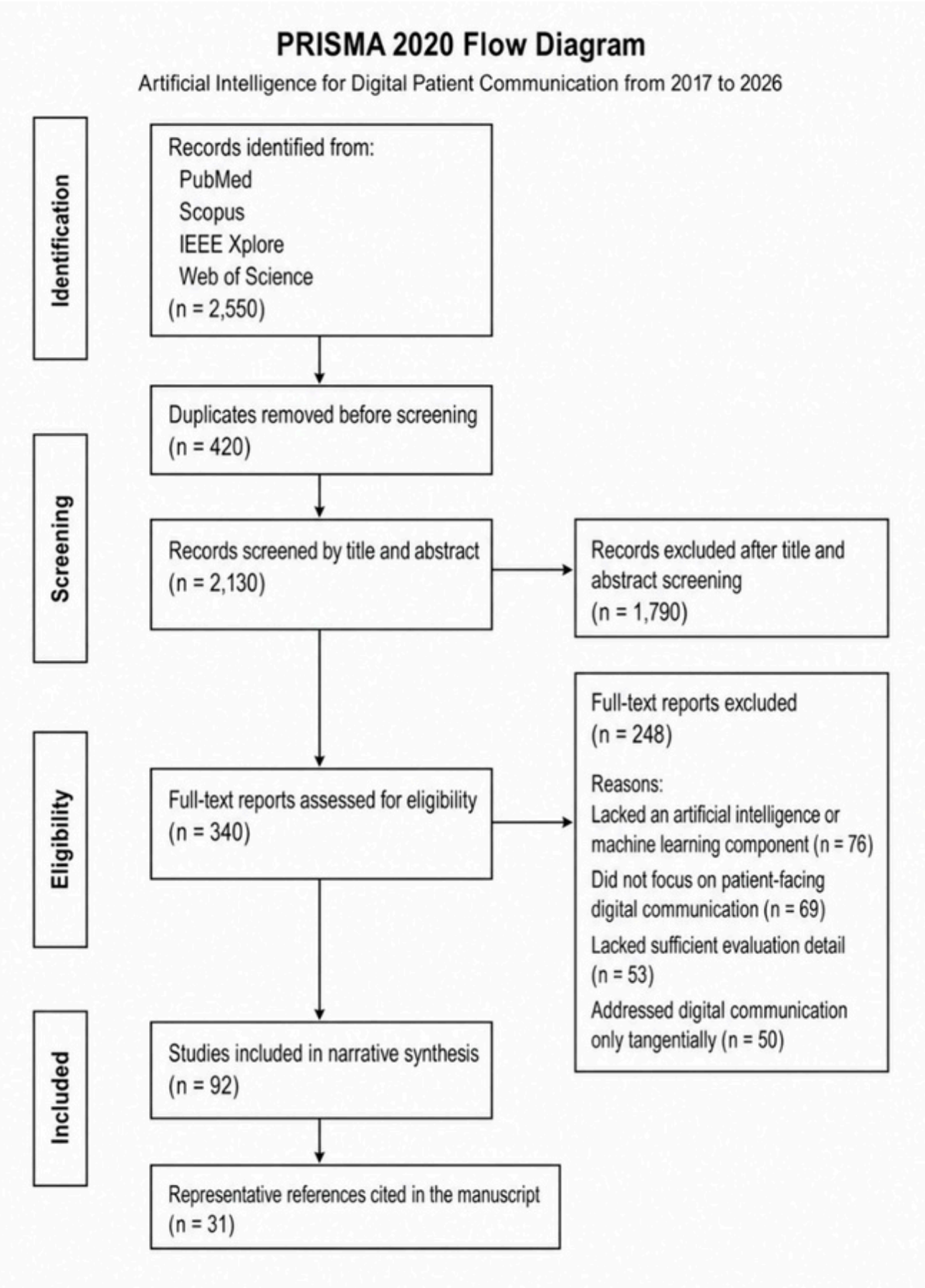
Inclusion and exclusion criteria

Eligible studies included original research, systematic reviews, scoping reviews, rapid reviews, or evaluation studies that examined AI-enabled tools for patient-facing digital communication in healthcare. Studies were included when they addressed portal message triage, chatbot or conversational-agent support, care navigation, automated response drafting, or analytics using digital communication data to assess engagement, satisfaction, adherence, or disengagement. Studies were excluded if they focused only on clinical-note NLP without patient-facing communication, general telehealth without an AI component, or purely technical model development without a healthcare communication context. English-language publications from 2017 to 2026 were eligible, and studies outside that time window were excluded.

Screening and selection

The search identified 2,550 records, of which 420 duplicates were removed before screening. Title and abstract screening excluded 1,790 records, leaving 340 full-text articles assessed for eligibility; of these, 248 were excluded because they lacked an AI component, did not focus on patient-facing communication, lacked sufficient evaluation detail, or addressed digital communication only tangentially. Ninety-two studies were included in the full narrative synthesis, with 31 references selected for direct citation in this manuscript to represent the major evidence domains and publication period. The selection process was designed to be consistent with PRISMA 2020 reporting expectations and to capture evidence from both mature domains, such as chatbots, and emerging domains, such as large language model-generated draft replies.

Figure 1 presents the PRISMA 2020 study-selection process from 2,550 identified records to 92 studies included in the narrative synthesis.



Patient Communication

Data extraction

Data were extracted using a structured form that captured study year, country or setting, patient population, communication channel, AI method, input data, task, comparator, and evaluation outcomes. Extracted technical details included whether systems used rule-based logic, classical machine learning, deep learning, transformer-based NLP, large language models, retrieval-augmented generation, or hybrid approaches [1, 8, 18]. Clinical and implementation details included whether the system was evaluated retrospectively, prospectively, in simulation, in live deployment, or through user-centered assessment. Safety extraction focused on escalation protocols, human review, harm reporting, equity considerations, and whether the tool had been integrated into EHR, portal, or workflow systems [15-17].

Risk of bias assessment

Risk of bias was assessed narratively because the included literature contained heterogeneous study designs, including model-development studies, retrospective analyses, surveys, feasibility studies, implementation evaluations, and reviews. For predictive or classification studies, the assessment drew on domains commonly emphasized in prediction-model appraisal, including participant selection, outcome definition, predictor measurement, validation strategy, and applicability to clinical workflow [1, 9, 19]. For chatbot and response-drafting studies, the assessment emphasized user selection, ecological validity, transparency of evaluation, human oversight, and the degree to which outcomes reflected real patient-clinician communication [11, 14, 20]. For implementation-oriented studies, attention was given to single-site design, vendor involvement, short follow-up, and whether workload, safety, and equity outcomes were independently assessed [7, 15, 21].

Synthesis methods

A narrative synthesis was conducted because the included studies varied substantially in AI methods, communication channels, populations, outcomes, and evaluation designs. Studies were grouped into five prespecified domains: portal message triage, chatbot support, care navigation, automated response drafting, and patient engagement analytics [4, 22, 23]. Within each domain, findings were summarized by task, model type, data source, evaluation approach, safety oversight, and implementation maturity. Cross-domain synthesis then compared recurring themes, including technical feasibility, limited real-world effectiveness evidence, human-in-the-loop dependence, insufficient equity analysis, and the shift from narrow automation toward AI-supported communication partnership [16, 17, 24].

Results and Discussion

Study selection

The final evidence base included 92 studies, with 31 representative references cited in this manuscript. The largest groups addressed chatbot support, conversational agents, and portal message analysis, while smaller but rapidly growing groups addressed automated response drafting and patient engagement analytics [4, 10, 11, 22]. Full-text exclusions most commonly reflected absence of patient-facing communication, lack of AI or machine learning methods, or evaluation of general digital health platforms without communication-specific outcomes. **Figure 1** should present the PRISMA flow from 2,550 identified records to 92 included studies, with exclusions at duplicate removal, title and abstract screening, full-text review, and final inclusion.

Study characteristics

Included studies were concentrated in the United States, Europe, and digitally mature health systems with established patient portals, EHR infrastructure, or consumer-facing chatbot deployments. Early studies from 2017 to 2020

emphasized portal-message classification, secure messaging patterns, and systematic characterization of conversational agents [1-6, 10, 11]. Studies from 2021 onward increasingly examined NLP-enabled message analysis, patient-clinician communication content, digital engagement phenotyping, and large language model-generated responses [8, 9, 14, 25]. The literature contained retrospective model-development studies, cross-sectional surveys, systematic reviews, implementation evaluations, and emerging real-world deployment reports.

Table 1 organizes the reviewed evidence into a functional communication architecture that distinguishes AI inputs, task logic, operational use, human oversight, and implementation maturity across the five major domains.

Table 1. Functional Architecture of Artificial Intelligence for Digital Patient Communication Across the Patient Communication Continuum

AI communication domain	Primary communication inputs	Core AI task logic	Typical operational output	Human role required	Evidence maturity pattern	Main implement risk
Portal message triage	Patient portal messages, secure inbox messages, message metadata, care-team routing information, urgency cues	Classifies message topic, intent, urgency, routing destination, or escalation priority using NLP, supervised learning, deep learning, transformer models, or LLM-supported prioritization	Routed inbox queue, urgency flag, specialty or staff assignment, prioritization label, emergency-detection alert	Clinicians or trained staff review high-risk classifications, validate routing, and manage ambiguous or urgent messages	Relatively mature compared with other domains; evidence includes early NLP classifiers, retrospective studies, and emerging prioritization approaches	False negative for urgent messages; oversimplified topic labels; poor handling of mixed clinical and administrative requests; routing accountability gaps
Chatbot support	Patient questions, symptom descriptions, scripted interactions, chatbot transcripts, health education queries, administrative requests	Uses rule-based decision trees, scripted conversational flows, retrieval-based responses, machine learning, or adaptive conversational-agent logic	Routine question answering, symptom guidance, education, administrative support, self-service interaction, or escalation recommendation	Human escalation remains necessary for uncertain, severe, sensitive, or clinically complex situations	Broadly studied but uneven; many studies are short-term, feasibility-oriented, or controlled rather than long-term and outcome-based	Inaccurate advice, with escalation design, limited empathy, variable clinical integration, insufficient surveillance
Care navigation AI	Digital front-door queries, scheduling requests, referral questions,	Interprets patient needs and maps them to next-step services, scheduling	Navigation recommendation, scheduling pathway, referral direction, follow-up prompt,	Staff oversight is needed when access barriers, referral complexity, clinical risk, or	Less mature as a distinct evidence domain; often embedded within chatbot,	Generic access without operational linkage; inequitable access

	follow-up needs, test-result questions, appointment access information	pathways, referral support, follow-up options, or care-team routing	service-line routing, or staff handoff	social needs are present	portal, or digital front-door systems	assumption of weak integration with scheduled and referral systems, unclear handoff responsibilities
Automated response drafting	Portal inbox messages, patient questions, EHR context, prior visit information, medications, test results, local protocols, clinician style expectations	Generates draft patient-facing replies using LLMs, transformer-based models, retrieval-augmented generation, or context-aware response-generation systems	Draft response for clinician review, edited patient reply, suggested phrasing, explanation text, or templated communication	Clinician review, editing, approval, and accountability are essential safety requirements	Rapidly emerging after LLM adoption; evidence includes simulation, clinician ratings, patient preference studies, and early EHR-integrated implementation	Hallucinations, inaccurate personalization, inappropriate tone, omission of clinical context, overreliance on data, and uncertainty about disclosure
Patient engagement analytics	Portal use patterns, message frequency, digital program activity, telehealth engagement, communication content, adherence traces, disengagement signals	Detects engagement phenotypes, predicts disengagement risk, estimates activation or adherence, or identifies patients needing outreach	Engagement-risk score, outreach trigger, coaching prompt, follow-up recommendation, or population-level engagement dashboard	Care teams must interpret predictions and decide whether outreach, education, care coordination, or escalation is appropriate	Underdeveloped and methodologically varied; fewer studies link predictions to tested interventions	Mislabeling of patients, disengagement when barriers reflect access language, increased caregiver burden, digital exclusion
Cross-domain integrated communication pathway	Combined portal, chatbot, navigation, drafting, and engagement data streams linked to EHR and operational systems	Coordinates classification, response support, navigation, escalation, and outreach across a continuous patient communication workflow	Accountable AI-supported communication infrastructure with audit trails, safety thresholds, human review, and measurable outcomes	Multidisciplinary oversight involving clinicians, nurses, administrative staff, informatics teams, governance leaders, and patient representatives	Largely absent from current evidence; no fully integrated end-to-end platform was evaluated across all five domains	Cascading errors, duplicated work, unclear accountability, fragmented governance, patient confusion, unmeasured workload redistribution

Portal message triage

Portal message triage studies applied NLP, machine learning, and deep learning to classify incoming patient messages by topic, intent, urgency, or routing destination. Earlier work demonstrated that convolutional neural networks and rule-based or machine-learning classifiers could categorize portal messages into clinically meaningful groups, supporting potential triage and routing workflows [1, 2]. Secure messaging analyses also showed that messages often combine administrative, informational, and clinical content, making triage more complex than simple topic labeling [3, 9]. More recent work extended this direction toward prioritization workflows and emergency detection, including large language model and retrieval-augmented approaches for identifying potentially urgent portal messages [19, 18].

Chatbot support

Chatbot support was one of the most developed areas in the literature, encompassing symptom checkers, conversational agents for health education, administrative support, and mental health-related engagement. Systematic and scoping reviews described a wide range of rule-based, retrieval-based, and machine-learning-enabled conversational agents used for patient interaction, although many tools were evaluated in short-term or controlled settings [4, 10, 11]. Studies of personalization and physician perceptions highlighted that patients and clinicians may value convenience and availability but remain concerned about accuracy, safety, empathy, and appropriate escalation [5, 13]. Recent reviews continued to find substantial enthusiasm for healthcare chatbots, while also emphasizing variability in evidence quality, clinical integration, and outcome assessment [24, 26].

Care navigation AI

Care navigation AI was less frequently studied as a distinct domain but appeared across digital front-door, referral guidance, follow-up support, and administrative communication use cases. The reviewed literature suggested that navigation tools often overlap with chatbot and portal functions, because patients commonly seek help with appointment scheduling, care access, test follow-up, referrals, billing, and next steps after clinical encounters [4, 10, 26]. Digital front-door approaches were framed as systems that could guide patients to appropriate services, reduce friction in access, and support routing before or after clinician involvement [27]. However, evidence for end-to-end AI navigation across referral management, longitudinal follow-up, and complex care coordination remained limited compared with evidence for isolated chatbot or triage tasks.

Automated response drafting

Automated response drafting emerged as a major growth area after the introduction of large language models into clinical communication workflows. Studies evaluating AI-generated draft replies to patient inbox messages reported that draft generation could support clinician communication, but consistently required human review because accuracy, tone, specificity, and safety depend on context [14-16]. Implementation-oriented work suggested that integrating draft replies into health records may change clinician workflow, communication style, and perceptions of message burden, although evidence on net workload reduction remained early and context-dependent [15, 21]. Pediatric and specialty-specific studies further showed that response drafting must account for population, caregiver involvement, clinical risk, and local protocols [28].

Patient engagement analytics

Patient engagement analytics used digital traces such as message frequency, portal behavior, communication content, or telehealth participation to estimate engagement, activation, adherence, satisfaction, or disengagement risk. Compared with chatbots and message triage, this domain was less common and more methodologically varied [23, 29]. Digital diabetes prevention and tele-mental health studies demonstrated that machine learning can be used to derive engagement phenotypes or estimate patient engagement from communication and session data [23, 29]. However, the

evidence base remained underdeveloped regarding how engagement predictions should trigger interventions, how patients should be informed, and whether such analytics improve outcomes.

AI methods used

The reviewed studies used a broad range of AI methods, from rule-based logic and classical supervised learning to convolutional neural networks, topic modeling, transformer-based NLP, and large language models. Earlier portal-message studies relied on structured labels, manually annotated corpora, and supervised classification, while later work increasingly used richer message representations and deep learning approaches [1, 2, 8]. Chatbot studies included rule-based decision trees, scripted conversational flows, retrieval mechanisms, and more adaptive conversational-agent architectures [4, 5, 11]. The newest response-drafting and emergency-detection studies reflected a shift toward generative AI and retrieval-augmented methods, raising new questions about hallucination, local grounding, and validation [14, 18].

Data sources and integration

Data sources included patient portal messages, provider-to-patient secure messages, chatbot transcripts, public patient questions, EHR-linked message records, and digital health program engagement logs. Portal studies showed that secure messages contain heterogeneous requests and often require linkage to clinical context, visit history, medications, test results, and care-team roles for accurate interpretation [3, 9, 25]. Chatbot and conversational-agent studies frequently relied on interaction logs, survey data, or simulated patient queries rather than full integration with clinical records [10, 13, 20]. Response-drafting studies placed greater emphasis on integration with the EHR inbox, because draft quality and safe clinician review depend on access to relevant patient context and workflow positioning [14, 15].

Outcome measures

Outcome measures varied substantially across domains and were often limited to technical or proximal endpoints. Portal triage studies commonly reported classification accuracy, routing performance, or prioritization feasibility, while chatbot studies frequently assessed satisfaction, perceived usefulness, engagement, or conversational quality [1, 4, 11]. Response-drafting studies examined clinician ratings, patient preferences, adoption, or utility of generated replies, but fewer studies directly measured downstream outcomes such as response time, clinician workload, safety events, or patient health status [14–17]. Overall, the evidence suggested that AI communication tools have been more consistently evaluated for feasibility and acceptability than for clinical effectiveness.

Safety and human oversight

Safety and human oversight were recurring concerns, particularly for tools that might influence triage, advice, escalation, or clinician-authored responses. Studies of emergency detection and portal prioritization underscored the need to identify high-risk messages reliably and to prevent urgent concerns from being misclassified as routine [19, 18]. Chatbot reviews similarly emphasized that conversational agents require escalation pathways, boundaries of use, and careful handling of mental health, symptom severity, and uncertainty [4, 11, 24]. Response-drafting studies consistently positioned AI as assistive rather than autonomous, with clinician review serving as the central safeguard against inaccurate, incomplete, or inappropriate replies [14–16].

Health equity considerations

Health equity was addressed inconsistently across the evidence base, despite the importance of language, health literacy, digital access, cultural adaptation, and algorithmic bias in patient communication. Reviews of conversational agents noted that personalization, accessibility, and user adaptation were often discussed but not rigorously evaluated across diverse populations [5, 10]. Studies of patient preferences and ethics in AI-drafted responses highlighted concerns about transparency, trust, therapeutic relationship, and the acceptability of AI involvement in sensitive clinical

communication [17]. Digital front-door and triage applications raised additional equity concerns because errors in routing, language interpretation, or access assumptions could disproportionately affect patients with lower digital literacy or limited portal access [27].

Implementation and deployment maturity

Implementation maturity ranged from retrospective model development to live EHR-integrated pilots and early deployment studies. Portal classification and secure-message analysis provided foundational evidence but often remained retrospective or exploratory, whereas recent response-drafting studies increasingly examined workflow integration and real clinician use [1, 3, 15]. Chatbot evidence included many prototypes and short-term evaluations, with fewer examples of sustained deployment linked to measurable care outcomes [11, 24, 26]. Overall, the field appears to be transitioning from proof-of-concept models toward operational AI communication tools, but long-term sustainability, governance, and independent replication remain limited [21, 30].

Chatbots and triage models are the most proliferated

Chatbots and portal-message triage models represented the most frequently studied applications of AI for digital patient communication. This pattern likely reflects the availability of message data, the operational pressure created by high communication volumes, and the relative feasibility of evaluating classification or conversational interactions in bounded settings [1, 4, 11]. Several studies reported technically promising results for message classification, conversational-agent support, and user-facing interaction, but fewer demonstrated durable improvements in clinical outcomes or system workload [2, 10, 24]. The evidence therefore supports technical feasibility more strongly than it supports broad real-world impact.

Automated response drafting is emerging, enabled by LLMs

Automated response drafting has grown rapidly as large language models have become capable of producing fluent, patient-facing text. Studies of AI-generated draft replies suggested that such tools may reduce repetitive composition work and improve response consistency, but they also create risks related to factual accuracy, personalization, tone, and clinician accountability [14–16]. Human-in-the-loop review was treated as essential across this literature, not merely as an implementation preference but as a safety condition for clinical communication. The emerging evidence indicates that generative AI may be useful as a drafting assistant, while fully autonomous patient communication remains unsupported.

Care navigation tools bridge communication and action

Care navigation tools occupy an important middle ground between information exchange and concrete care coordination. The reviewed evidence suggested that navigation functions are often embedded within chatbots, digital front doors, or portal workflows rather than studied as independent AI systems [10, 26, 27]. For these tools to be clinically useful, they must connect communication with scheduling, referrals, follow-up, and care-team routing rather than simply provide generic advice. Integration with EHRs, referral systems, and organizational protocols therefore appears central to the future of AI-enabled care navigation.

Patient engagement analytics are under-commercialized

Patient engagement analytics appeared less mature than chatbot, triage, or draft-response applications, despite their strategic relevance for proactive care. Studies using digital diabetes prevention and tele-mental health data showed that machine learning can identify engagement patterns or estimate engagement intensity from digital behavior and communication traces [23, 29]. However, few studies demonstrated how engagement predictions should be operationalized into outreach, coaching, clinical escalation, or shared decision-making. This gap suggests that engagement analytics may have substantial potential but requires stronger linkage between prediction and actionable intervention.

Safety, equity, and language remain sidelined

Safety, equity, and language adaptation were often acknowledged but rarely evaluated with the same rigor as technical performance or usability. Chatbot and conversational-agent reviews repeatedly identified concerns about escalation, misinformation, empathy, and vulnerable populations, yet many primary studies lacked detailed harm surveillance or subgroup analysis [4, 5, 11]. AI-drafted response studies raised additional concerns about transparency and patient trust, especially when patients may not know whether a response was AI-assisted [17]. Because digital communication is shaped by literacy, language, access, and trust, limited equity evaluation represents a major weakness in the current evidence base.

The human-AI communication partnership

The reviewed literature consistently supported a partnership model in which AI augments rather than replaces human communication. Portal triage models may help sort or prioritize messages, chatbots may answer routine questions or collect structured information, and draft-reply systems may accelerate clinician response, but each function requires boundaries and escalation rules [14-16, 19]. The optimal handoff between AI and human clinicians remains unresolved, particularly when patient concerns are ambiguous, emotionally sensitive, or clinically urgent. Future communication systems will need to define when AI should classify, suggest, draft, escalate, or remain silent.

Figure 2 synthesizes the review findings into an evidence-to-implementation map linking digital communication inputs, AI method families, operational use pathways, safety requirements, human oversight, and future research priorities.

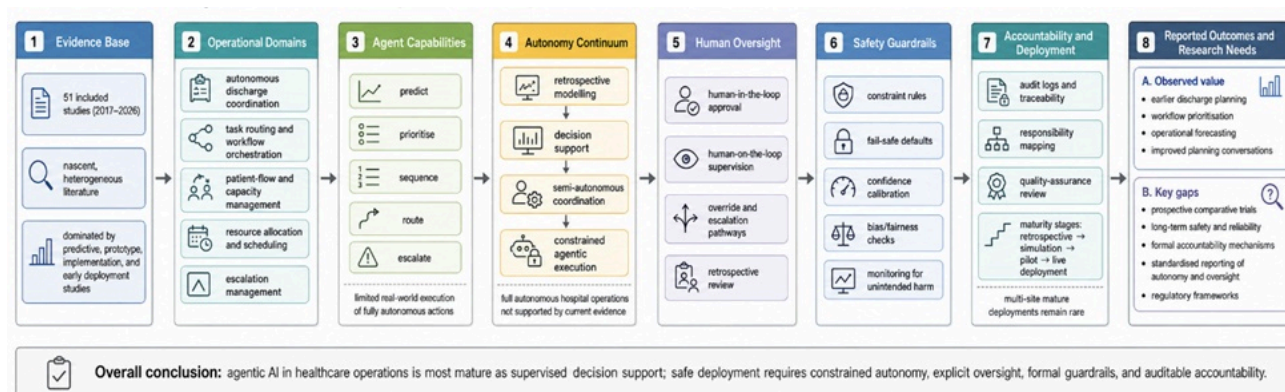


Figure 2. Evidence-to-Implementation Map of Artificial Intelligence for Digital Patient Communication

Evaluation gap: from accuracy to outcomes

A major cross-cutting finding was that evaluation often stopped at model performance, user satisfaction, or perceived utility. Although such endpoints are important, they do not establish whether AI communication tools reduce clinician burden, improve timeliness, enhance patient understanding, or prevent harm [6, 7, 15]. Studies of patient-message volume and response-drafting implementation suggest that workflow outcomes must be measured directly, because introducing AI can redistribute work rather than simply reduce it [7, 21]. The field therefore needs stronger prospective studies that connect AI-supported communication to patient-level, clinician-level, and system-level outcomes.

Limitations

Review limitations

This review has several limitations related to scope, search strategy, and evidence heterogeneity. Restricting the synthesis to English-language peer-reviewed publications may have excluded relevant studies of multilingual chatbots, regional digital front-door systems, or non-English patient communication tools [5, 10]. The broad definition of AI for digital

patient communication brought together diverse methods and settings, which supported conceptual synthesis but limited direct comparison across outcomes [4, 24]. In addition, rapid development of large language models means that studies published near the end of the review period may already reflect evolving tools, workflows, and governance expectations [14, 18].

Evidence base limitations

The underlying evidence base also had important limitations. Many studies were retrospective, single-site, simulation-based, or focused on technical feasibility rather than independent prospective evaluation in routine care [1, 15, 19]. Chatbot and conversational-agent studies often had short follow-up periods, variable reporting of safety processes, and limited assessment of equity, language, or access barriers [11, 24, 26]. Response-drafting and engagement-analytics studies were promising but still early, with few independent replications, few long-term workload assessments, and limited evidence on whether AI-supported communication improves patient outcomes [21, 23, 29, 30].

Comparison with prior reviews

Prior reviews in this area have generally focused on narrower categories, particularly conversational agents, chatbot effectiveness, personalization, or chronic-condition support. These reviews established that chatbots and conversational agents are widely studied in healthcare, but they often treated patient communication as a chatbot-specific phenomenon rather than as part of a broader digital communication ecosystem [4, 5, 10, 11]. Other related work examined patient portal content, secure-message classification, or NLP methods for message analysis, but did not systematically connect these methods to care navigation, response drafting, or engagement analytics [1, 2, 8, 9]. As a result, prior syntheses provided important domain-specific insight while leaving the full patient communication pipeline only partially described.

This review extends earlier work by synthesizing AI applications across the communication continuum, from message intake and triage to chatbot support, navigation, response drafting, and engagement analysis. The inclusion of newer studies on AI-generated draft replies and portal-message prioritization captures a shift from standalone tools toward AI embedded in clinical inboxes and EHR workflows [14-16, 19]. Similarly, the inclusion of digital engagement phenotyping and tele-mental health engagement estimation highlights how communication data can support proactive care rather than only reactive message handling [23, 29]. This broader framing shows that AI for patient communication is not a single technology category but a set of linked capabilities that may shape access, workload, and patient experience.

The review also highlights safety and equity dimensions that have often been secondary in prior chatbot-centered syntheses. Studies of physician perceptions, patient preferences, and ethics in AI-drafted replies suggest that trust, transparency, therapeutic relationship, and acceptability are central to patient-facing AI communication [13, 17]. Work on digital front doors further indicates that navigation tools may amplify or reduce inequities depending on language support, health literacy accommodation, escalation design, and access assumptions [27]. By comparing these themes across domains, the review suggests that future evaluations should treat safety, equity, and communication quality as core outcomes rather than peripheral implementation concerns.

Recommendations

Evaluation standards

Future studies should use evaluation standards that extend beyond technical performance and short-term user satisfaction. Portal triage systems should report not only classification performance but also routing consequences, escalation safety, false reassurance risk, and impact on response timeliness [19, 18]. Chatbot and conversational-agent studies should include predefined safety outcomes, escalation success, patient comprehension, and subgroup analyses for language, literacy, and digital access [5, 11, 24]. Response-drafting evaluations should measure clinician editing burden, message accuracy, tone, patient trust, and downstream workload rather than only perceived usefulness [14-16].

Table 2 translates the review findings into an evaluation framework for moving AI-supported patient communication from technical feasibility toward safety, equity, workload impact, and accountable implementation.

Table 2. Evidence Gaps and Evaluation Priorities for Safe, Equitable, and Clinically Useful AI-Supported Patient Communication

Evaluation priority	Why it matters for digital patient communication	Minimum evidence currently needed	Recommended measures	Best-fit study design	Governance implication
Urgency and safety detection	Patient messages may contain urgent symptoms, medication concerns, mental health risks, or deterioration signals that must not be misclassified as routine	Evidence on false negatives, missed escalation, delayed response, and unsafe reassurance across real portal and chatbot workflows	Sensitivity for urgent messages, false-negative rate, escalation timeliness, adverse communication events, manual override frequency	Prospective silent trial followed by monitored deployment with predefined safety thresholds	Urgent-message AI should require conservative thresholds, rapid escalation pathways, audit trails, and human review
Communication quality	AI-generated or AI-mediated communication can affect patient understanding, trust, reassurance, and therapeutic relationship	Evidence that AI-supported replies are accurate, understandable, empathetic, and appropriate for the patient's context	Accuracy, tone, readability, empathy ratings, patient comprehension, clinician editing burden, patient trust, disclosure acceptability	Randomized vignette studies, clinician-review studies, patient-preference studies, and pragmatic implementation studies	AI-generated replies should remain reviewable, editable, attributable, and transparent to patients where disclosure is required
Workflow and workload impact	AI tools may redistribute work rather than reduce it, especially if clinicians must correct, verify, or monitor outputs	Direct measurement of clinician time, inbox burden, after-hours work, routing efficiency, and staff task allocation	Response time, message volume per role, after-hours work, time-to-resolution, number of touches per message, clinician editing time	Pre-post implementation studies, interrupted time-series analysis, pragmatic trials, and workflow observation	Deployment should be judged as an operational intervention, not only as a model or interface
Equity and access	Digital communication tools may perform differently by language, literacy, disability, socioeconomic status, age, cultural context, or portal access	Subgroup performance evidence and evaluation of whether AI improves or worsens access for digitally underserved groups	Language-specific performance, readability, subgroup error rates, portal access stratification, differential escalation, patient-reported barriers	Stratified validation studies, equity audits, multilingual usability testing, and community-informed evaluation	Equity monitoring should be built into model validation, procurement, deployment, and post-market surveillance

Human-AI handoff quality	The most important safety boundary is often the transition from AI suggestion to human judgment, escalation, or approval	Evidence defining when AI should classify, draft, answer, escalate, defer, or remain silent	Handoff completion rate, override rate, escalation appropriateness, clinician agreement, ambiguity handling, unresolved-message rate	Mixed-methods implementation evaluation with workflow mapping and prospective monitoring	Governance must specify role boundaries, accountability, escalation rules, and documentation of human review
Clinical and patient-level outcomes	Technical success does not prove that AI communication improves care quality or patient experience	Evidence linking AI communication tools to meaningful outcomes beyond feasibility or satisfaction	Care delay, follow-up completion, referral completion, medication clarification, avoidable urgent visits, patient experience, health status when relevant	Pragmatic randomized trials, controlled implementation studies, and longitudinal cohort evaluations	AI communication systems should not be considered mature until they demonstrate patient-centered and system-level benefit
Transparency and trust	Patients may respond differently when communication is AI-assisted, especially for sensitive topics or clinician-authored replies	Evidence on patient preferences, disclosure expectations, trust, and acceptability across clinical contexts	Disclosure preference, perceived authenticity, trust in clinician, comfort with AI involvement, preference by message type	Patient survey experiments, qualitative studies, ethics-informed implementation evaluations	Health systems should define when and how AI involvement is disclosed in patient-facing communication
Long-term sustainability	Short pilots may miss model drift, changing communication volume, staff adaptation, and evolving patient expectations	Evidence on durability, monitoring, maintenance, retraining, governance costs, and independent replication	Model drift, sustained workload impact, safety event trends, adoption persistence, staff burden, maintenance effort	Longitudinal deployment studies, registry-based monitoring, and multicenter replication	AI communication tools require lifecycle governance, not one-time validation

Human oversight and governance

Health systems implementing AI for digital patient communication should define explicit human oversight models before deployment. For triage tools, this includes thresholds for urgent escalation, review queues, audit trails, and accountability for routing failures [18, 19]. For chatbots and navigation systems, governance should specify what the tool may answer, when it must defer, and how patients are connected to human staff when needs exceed automated support [4, 26, 27]. For AI-generated draft replies, clinician review should remain mandatory unless rigorous evidence demonstrates safety for a narrowly defined low-risk communication use case [14, 15, 17].

Integration with clinical workflows

AI communication tools should be evaluated as workflow interventions, not only as algorithms or interfaces. Secure-message studies show that patient communication requires coordination across clinicians, nurses, administrative staff, and institutional policies, meaning that automation may shift work rather than eliminate it [3, 6, 7]. Integration with EHR data, scheduling systems, referral pathways, and care-team assignments is especially important for tools that aim to support navigation or generate context-sensitive replies [15, 25, 27]. Implementation studies should therefore examine how AI changes task allocation, clinician attention, patient expectations, documentation practices, and after-hours work.

Research gaps

Integrated communication platforms

No included study evaluated a fully integrated end-to-end AI communication platform that combined portal triage, chatbot support, care navigation, automated response drafting, and patient engagement analytics in one continuous workflow. Existing studies typically examined one function at a time, such as message classification, conversational support, draft reply generation, or engagement estimation [1, 4, 14, 23]. This fragmented evidence base makes it difficult to determine whether combining tools would improve continuity or create new risks through cascading errors, duplicated work, or unclear accountability. Future research should evaluate integrated platforms with explicit attention to handoffs, escalation, auditability, and patient understanding of AI involvement.

Rigorous safety and efficacy trials

The evidence base lacks rigorous trials assessing whether AI-driven patient communication improves safety, care outcomes, clinician workload, or patient experience under real-world conditions. Many studies used retrospective data, simulated questions, cross-sectional perceptions, or early implementation measures rather than randomized or controlled prospective designs [13, 15, 19, 20]. For high-stakes functions such as urgency detection, symptom guidance, mental health support, and draft clinical replies, stronger evidence is needed on false negatives, inappropriate reassurance, delayed care, and clinician overreliance [11, 18, 24]. Future trials should include patient-level outcomes, workload measures, safety monitoring, and predefined stopping criteria for communication-related harms.

Digital communication equity

Digital communication equity remains a major research gap because relatively few studies tested AI performance across language, literacy, disability, race, ethnicity, socioeconomic status, or portal access patterns. Personalization and patient preference studies suggest that users may differ in how they perceive conversational agents and AI-drafted replies, but subgroup evidence remains limited [5, 13, 17]. Engagement analytics may also risk labeling patients as disengaged when the underlying issue is access, trust, language, caregiving burden, or digital exclusion [23, 29]. Future research should evaluate whether AI communication tools reduce disparities by improving access and navigation or worsen them by privileging patients who are already digitally connected.

Conclusion

AI for digital patient communication is a dynamic field with strong technical achievements in chatbots, portal message triage, and automated response drafting. The evidence suggests that these tools can classify messages, support routine interactions, and assist clinicians with response composition in ways that may become increasingly important as digital communication volumes grow.

Care navigation and patient engagement analytics are less mature but strategically vital for patient-centered care. Their greatest value may come from connecting communication to action, including referral completion, follow-up support, proactive outreach, and earlier recognition of disengagement.

The literature remains characterized by a strong focus on technical performance, feasibility, and usability, with a concerning lack of rigorous safety, equity, and outcome-oriented studies. More evidence is needed to determine whether

AI-supported communication improves care quality, reduces workload, protects patient trust, and serves diverse populations fairly.

To realize the promise of AI-augmented communication, future research must prioritize patient-level impact, safety, transparency, and equitable deployment. The next phase of the field should move from isolated tools toward accountable, integrated, human-centered communication systems.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 13 Jan 2026

Revised: 15 Feb 2026

Accepted: 15 Apr 2026

Published online: 20 July 2026

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at <https://creativecommons.org/licenses/by-nc-sa/4.0/>.

References

- Suliman L, Gilmore D, French C, Cronin RM, Jackson GP, Russell M, et al. Classifying patient portal messages using convolutional neural networks. *J Biomed Inform.* 2017;74:59-70.
- Cronin RM, Fabbri D, Denny JC, Rosenbloom ST, Jackson GP. A comparison of rule-based and machine learning approaches for classifying patient portal messages. *Int J Med Inform.* 2017;105:110-20.
- Shimada SL, Petrakis BA, Rothendler JA, Zirkle M, Zhao S, Feng H, et al. An analysis of patient-provider secure messaging at two Veterans Health Administration medical centers: message content and resolution through secure messaging. *J Am Med Inform Assoc.* 2017;24(5):942-9.

Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. *J Am Med Inform Assoc.* 2018;25(9):1248-58.

Kocaballi AB, Berkovsky S, Quiroz JC, Laranjo L, Tong HL, Rezazadegan D, et al. The personalization of conversational agents in health care: systematic review. *J Med Internet Res.* 2019;21(11):e15360.

North F, Luhman KE, Mallmann EA, Mallmann TJ, Tulledge-Scheitel SM, North EJ, et al. A retrospective analysis of provider-to-patient secure messages: how much are they increasing, who is doing the work, and is the work happening after hours? *JMIR Med Inform.* 2020;8(7):e16521.

Holmgren AJ, Byron ME, Grouse CK, Adler-Milstein J. Association between billing patient portal messages as e-visits and patient messaging volume. *JAMA.* 2023;329(4):339-42.

TaftiAhmad P. Probing patient messages enhanced by natural language processing: a top-down message corpus analysis. *Health Data Sci.* 2021;2021:1-8.

Huang M, Fan J, Prigge J, Shah ND, Costello BA, Yao L. Characterizing patient-clinician communication in secure medical messages: retrospective study. *J Med Internet Res.* 2022;24(1):e17273.

Tudor Car L, Dhinakaran DA, Kyaw BM, Kowatsch T, Joty S, Theng YL, et al. Conversational agents in health care: scoping review and conceptual analysis. *J Med Internet Res.* 2020;22(8):e17158.

Milne-Ives M, De Cock C, Lim E, Shehadeh MH, De Pennington N, Mole G, et al. The effectiveness of artificial intelligence conversational agents in health care: systematic review. *J Med Internet Res.* 2020;22(10):e20346.

Schachner T, Keller R, von Wangenheim F. Artificial intelligence-based conversational agents for chronic conditions: systematic literature review. *J Med Internet Res.* 2020;22(9):e20701.

Palanica A, Flaschner P, Thommandram A, Li M, Fossat Y. Physicians' perceptions of chatbots in health care: cross-sectional web-based survey. *J Med Internet Res.* 2019;21(4):e12887.

Garcia P, Ma SP, Shah S, Smith M, Jeong Y, Devon-Sand A, et al. Artificial intelligence-generated draft replies to patient inbox messages. *JAMA Netw Open.* 2024;7(3):e243201.

Tai-Seale M, Baxter SL, Vaida F, Walker A, Sitapati AM, Osborne C, et al. AI-generated draft replies integrated into health records and physicians' electronic communication. *JAMA Netw Open.* 2024;7(4):e246565.

English E, Laughlin J, Sippel J, DeCamp M, Lin CT. Utility of artificial intelligence-generative draft replies to patient messages. *JAMA Netw Open.* 2024;7(10):e2438573.

Cavalier JS, Goldstein BA, Ravitsky V, Bélisle-Pipon JC, Bedoya A, Maddocks J, et al. Ethics in patient preferences for artificial intelligence-drafted responses to electronic messages. *JAMA Netw Open.* 2025;8(3):e250449.

Liu S, Wright AP, McCoy AB, Huang SS, Steitz B, Wright A. Detecting emergencies in patient portal messages using large language models and knowledge graph-based retrieval-augmented generation. *J Am Med Inform Assoc.* 2025;32(6):1032-9.

Yang J, So J, Zhang H, Jones S, Connolly DM, Golding C, et al. Development and evaluation of an artificial intelligence-based workflow for the prioritization of patient portal messages. *JAMIA Open.* 2024;7(3):oae078.

Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA Intern Med.* 2023;183(6):589-96.

Bootsma-Robroeks CM, Workum JD, Schuit SC, Hoekman A, Mehri T, Doornberg JN, et al. AI-generated draft replies to patient messages: exploring effects of implementation. *Front Digit Health.* 2025;7:1588143.

Aggarwal A, Tam CC, Wu D, Li X, Qiao S. Artificial intelligence–based chatbots for promoting health behavioral changes: systematic review. *J Med Internet Res.* 2023;25:e40789.

Rodriguez DV, Chen J, Viswanadham RV, Lawrence K, Mann D. Leveraging machine learning to develop digital engagement phenotypes of users in a digital diabetes prevention program: evaluation study. *JMIR AI.* 2024;3:e47122.

Kurniawan MH, Handiyani H, Nuraini T, Hariyati RT, Sutrisno S. A systematic review of artificial intelligence-powered chatbot intervention for managing chronic illness. *Ann Med.* 2024;56(1):2302980.

De A, Huang M, Feng T, Yue X, Yao L. Analyzing patient secure messages using a FHIR-based data model: development and topic modeling study. *J Med Internet Res.* 2021;23(7):e26770.

Laymouna M, Ma Y, Lessard D, Schuster T, Engler K, Lebouché B. Roles, users, benefits, and limitations of chatbots in health care: rapid review. *J Med Internet Res.* 2024;26:e56930.

Alamoudi A, Kontopantelis E, Zghebi S, Brown B. AI triage in primary care: building safer and more equitable real-world evidence (preprint). *J Med Internet Res.* 2025;27:e12345.

Liang AS, Vedak S, Dussaq A, Yao DH, Villarreal JA, Thomas S, et al. Artificial intelligence–generated draft replies to patient messages in pediatrics. *JAMIA Open.* 2025;8(6):ooaf159.

Guhan P, Awasthi N, McDonald K, Bussell K, Reeves G, Manocha D, et al. Developing a machine learning–based automated patient engagement estimator for telehealth: algorithm development and validation study. *JMIR Form Res.* 2025;9:e46390.

Mandal S, Wiesenfeld BM, Szerencsy AC, Small WR, Major V, Richardson S, et al. Utilization of generative AI-drafted responses for managing patient-provider communication. *NPJ Digit Med.* 2025;8(1):591.