UNIVERSITY OF PAVIA – IUSS SCHOOL FOR ADVANCED STUDIES PAVIA

Department of Brain and Behavioral Sciences (DBBS)

MSc in Psychology, Neuroscience and Human Sciences

Perceived Empathy in Chatbot Interactions for Chronic Disease Management: A Scoping Review

How do users with chronic conditions perceive and experience empathy in chatbot-assisted disease management, and what factors influence these perceptions?

Supervisor:

Prof. Serena Barello

Co-Supervisor:

Michela Monaci, PhD Candidate

Thesis written by:

Narmin Alizada

Student ID: 549699

Academic Year 2025-2026

"Considerate la vostra semenza:
fatti non foste a viver come bruti,
ma per seguir virtute e canoscenza."

— Dante Alighieri

Contents

Abstract

Chronic disease management places sustained emotional and practical demands on patients, and healthcare systems are increasingly turning to chatbot-based interventions to extend support beyond the traditional clinical encounters. Yet the role of empathy in shaping patients' lived experience of these tools has received uneven attention: the design of empathic features has advanced faster than our understanding of how patients perceive and respond to them. This scoping review was conducted to map the current evidence on how users with chronic conditions perceive and experience empathy in chatbot-assisted disease management, and to examine the chatbot design strategies, user outcomes, and contextual factors that influence those perceptions.

Following Joanna Briggs Institute methodology and PRISMA-ScR reporting guidelines, a systematic search was conducted across five databases, namely PubMed, Scopus, Web of Science, PsycINFO, and CINAHL, with the final search update conducted on August 5, 2025. Eligibility criteria were defined using the Population, Concept, and Context framework. Studies were included if they involved users managing chronic or long-term health conditions, addressed perceived empathy or empathy-related relational constructs in chatbot interactions, and were situated within healthcare or self-management contexts. Forty studies met the criteria for inclusion and were synthesized using thematic and narrative synthesis.

The findings reveal that empathic design features, including personalized language, active listening cues, and emotionally responsive framing, are associated with more positive user perceptions of trust, satisfaction, and therapeutic alliance. However, the construct of digital empathy remains inconsistently operationalized across studies, with fewer than half of the included studies formally assessing empathy. User outcomes vary considerably depending on disease type, interaction modality, and the underlying chatbot architecture, with generative AI-powered systems generally associated with higher perceived empathy than earlier rule-based systems. The emergence of large language models has also begun to shift user expectations upward, such that chatbots regarded as adequate at the time of their evaluation may appear insufficient as users gain familiarity with more capable systems.

This review highlights both the genuine promise of empathic chatbot design and the methodological fragmentation that limits what can be concluded from the existing evidence. It identifies priority areas for future research, including standardized measurement instruments, theory-grounded design approaches, and greater representation of underserved populations. These findings carry practical implications for developers, clinicians, and policymakers seeking to integrate conversational agents into chronic disease care in a way that is both clinically effective and relationally meaningful.

Keywords: chatbot, conversational agent, empathy, chronic disease management, digital health, user experience, scoping review

Acknowledgements

I would like to express my biggest thanks to Professor Serena Barello and PhD Candidate Michela Monaci, my thesis supervisors and greatest supporters throughout my whole thesis journey. Thank you for all your encouragement, patience, dedication to helping, and for the feedback you provided at every stage of this work.

I would also like to thank my husband, whose support extended well beyond this thesis and accompanied me throughout my entire master's degree. He carried much of the mental and emotional weight of this process alongside me, listened through every difficult stretch, and helped me think through ideas when I could not see them clearly on my own. His willingness to share in both the intellectual and emotional load of this journey, from the very beginning of my studies to the final pages of this thesis, made a difference I cannot fully put into words, and I am deeply grateful to have had him beside me through it.

I also want to thank my family, and especially my mother, for her unwavering support, including in my decision to pursue this master's degree in the first place. Her encouragement and belief in me sustained me through the long process of completing this work, and her support has meant more to me than she may realize.

Finally, I am grateful to my friends for their support and for believing in me throughout this process.

Abbreviations

- **AI** – Artificial Intelligence
- **AIPL** – Artificial Intelligence Patient Librarian
- **CBT** – Cognitive Behavioral Therapy
- **COPD** – Chronic Obstructive Pulmonary Disease
- **DBT** – Dialectical Behavior Therapy
- **ECA** – Embodied Conversational Agent
- **EAL** – Empathic Active Listening
- **ESA** – Empathic Self-Awareness
- **GAI** – Generative Artificial Intelligence
- **GPT** – Generative Pre-trained Transformer
- **HIV** – Human Immunodeficiency Virus
- **IBD** – Inflammatory Bowel Disease
- **JBI** – Joanna Briggs Institute
- **MI** – Motivational Interviewing
- **mHealth** – Mobile Health
- **N** – Sample Size
- **NA** – Not Applicable / Not Available
- **PCC** – Population, Concept, and Context
- **PRISMA-ScR** – Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews

- **PTSD** – Post-Traumatic Stress Disorder
- **RCT** – Randomized Controlled Trial
- **SUD** – Substance Use Disorder
- **VTA** – Virtual Therapeutic Agent

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background: The Burden of Chronic Diseases and Digital Health Transition

Chronic diseases are, by any measure, among the defining health challenges of our time. Conditions such as type 2 diabetes, COPD, chronic liver disease, arthritis, HIV infection, and cancer affect billions of people worldwide and account for more than 70% of all deaths globally each year (MacNeill et al., 2024). Their toll is not only measured in mortality: chronic diseases are a leading cause of long-term disability and generate healthcare costs that strain systems across both high- and low-income settings. The scale of this burden becomes concrete when one considers specific conditions; the global prevalence of diabetes alone rose from 108 million in 1980 to 463 million by 2010 (Bruijnes et al., 2023), and projections suggest that one in six adults in Singapore will be affected by 2050 (Pienkowska et al., 2023). Beyond the physical challenges, the substantial psychological burden associated with these conditions, such as persistent anxiety and the feeling of solitary accountability, often represents a critical yet underrecognized challenge for patients (Easton et al., 2019; Pienkowska et al., 2023).

The psychological dimension of this burden runs far deeper than mortality and disability statistics convey. Living with a chronic illness is not simply a matter of managing physical symptoms; it almost inevitably involves confronting uncertainty, disruption to daily life, and very often significant emotional distress. Pienkowska et al. (2023), in a co-design study with adults living with type 2 diabetes, found that patients described their self-management experience in terms of solitary accountability, the difficulty of translating knowledge into practice, and an ongoing trial-and-error process of learning how their own body responds to dietary and lifestyle changes. These are not merely logistical challenges; they reflect the profound emotional and cognitive burden that chronic illness places on individuals who must navigate it largely on their own. Around 30% of people with one or more physical long-term conditions also experience a comorbid mental health disorder, most commonly anxiety or depression (Easton et al., 2019). This is particularly pronounced in conditions like COPD, where patients are up to ten times more likely to experience panic attacks than the general population, and where mental health comorbidities have a clinically significant impact on quality of life and healthcare utilisation (Easton et al., 2019; ter Stal et al., 2021). When depression or anxiety goes unaddressed alongside a physical illness, the consequences extend beyond emotional suffering: disease management becomes harder, medical outcomes worsen, and the overall burden on both patients and healthcare systems deepens (MacNeill et al., 2024). This emotional distress is further intensified by a widespread lack of access to professional mental health resources, which leaves many individuals to cope with their clinical and emotional needs without sufficient support (MacNeill et al., 2024).

Yet despite how common this overlap is, mental health support for people with chronic disease remains difficult to access. The barriers are well documented, as financial constraints, geographic distance from services, limited mobility, and a persistent shortage of trained providers all conspire to widen the gap between need and care (MacNeill et al., 2024). For many patients, treatment for the mental health dimension of their illness simply never arrives, and the gap between need and provision remains one of the most pressing structural failures in chronic disease care.

It is precisely this unmet need that has driven growing interest in digital health technologies as a complement to conventional care. The widespread availability of smartphones and internet connectivity has opened up possibilities for delivering health support in ways that were not previously feasible, at any hour, from any location, without a waiting list. Technology-based interventions for chronic conditions are generally more accessible and less costly than face-to-face services, and they offer a degree of anonymity that some patients find easier to engage with (Hauser-Ulrich et al., 2020). Virtual agents, in particular, are scalable and available around the clock, which makes them a plausible means of extending care beyond the walls of the clinic (van Bussel et al., 2022). Bruijnes et al. (2023) make this point directly in the context of diabetes, arguing that conversational agents can reach the millions of people with social diabetes distress who are currently not targeted or reached by traditional healthcare.

The range of chronic conditions that researchers have explored in this context is broad, spanning chronic pain and COPD, substance use disorders, HIV, cancer, and autoimmune diseases (Easton et al., 2019; Prochaska et al., 2021a, 2021b; Cook et al., 2024; Espinoza et al., 2025). Across this body of work, chatbots and conversational AI agents have been proposed not as replacements for human clinicians, but as what Boustani et al. (2021)

aptly describe as clinician extenders, tools that can reach people who would otherwise fall through the cracks of an already overburdened system.

Yet enthusiasm for these tools must be tempered by a clear-eyed look at how patients actually perceive them. Wu et al. (2024), in a survey of 888 adults with chronic conditions, found that fewer than one in three respondents would be likely to use a health chatbot in the next twelve months if one were available, with most expressing uncertainty not about functional capability but about whether a chatbot could respond to them in a way that felt sufficiently accurate and trustworthy. Trust, in their regression analysis, emerged as the strongest attitudinal predictor of chatbot adoption, outweighing privacy concerns and prior digital experience alike. What this tells us is that the uptake of conversational agents in chronic disease care is not primarily a question of technical availability. It is a question of how patients relate to these tools, whether they trust them, whether they feel genuinely heard by them, and whether the interaction carries enough relational weight to sustain engagement over time. These are the questions that motivate the review that follows, and it is with them in mind that the next section examines what conversational agents currently do in chronic disease management and what the evidence reveals about their effectiveness as support tools.

1.2 Conversational Agents as Support Tools in Chronic Disease Management

Conversational agents, broadly defined, are software systems designed to simulate human-like dialogue with users through text or voice, using a combination of scripted logic, natural language processing, and increasingly, artificial intelligence (Au et al., 2023; Rathnayaka et al., 2022). In healthcare, the term encompasses a wide spectrum of tools: from simple rule-based systems that guide users through predefined decision trees, such as the SingHealth RheumConnect chatbot used for symptom triage in rheumatology (Tan et al., 2023), to sophisticated AI-driven agents capable of open-ended conversation, such as the large language model-based GPT-4 evaluated in the context of oncology consultations (Bai et al., 2025). More recently, embodied conversational agents (ECAs), which present as animated, human-like avatars with facial expressions and nonverbal behaviours, have expanded the interactional possibilities still further, offering a form of multimodal engagement that goes beyond text exchange (Boustani et al., 2021; Martinez-Miranda et al., 2019). Whether disembodied or embodied, text-based or voice-enabled, these systems share a fundamental purpose: to engage users in a responsive, personalised dialogue that supports some dimension of their health. Throughout this introduction, the term "chatbot" is used as the general designator for text- and voice-based conversational software deployed in healthcare settings. The terms “conversational agent” and “virtual agent” are treated as broadly equivalent and used interchangeably, following the conventions of the studies cited. “Embodied conversational agent” (ECA) refers specifically to systems that present as animated avatars with visible facial expressions and gestural capabilities, and is distinguished from text-based chatbots where that distinction is experientially relevant.

The range of functions that conversational agents have been designed to serve in chronic disease management is considerable. At the most foundational level, many systems focus on patient education and health literacy, providing accessible, on-demand information about conditions, medications, and self-care strategies. Au et al. (2023) developed Lucy LiverBot specifically to address the gap in health literacy among patients with decompensated cirrhosis, a condition where cognitive impairment and complex care demands make information processing particularly difficult. Similarly, the AIPL chatbot evaluated by Leung et al. (2025) with metastatic breast cancer patients was designed to deliver conversational education about treatment options, symptom management, and cancer care resources. Beyond information provision, many agents are designed to support active self-management, providing reminders, tracking patient-reported symptoms, and offering structured feedback loops. The Brisa chatbot, studied by Cook et al. (2024) among adults with asthma, performed risk assessments and monitored symptom patterns over time, while the MARVIN system supported antiretroviral therapy adherence and self-management in people living with HIV through an AI-powered conversational interface on Facebook Messenger (Ma et al., 2025).

A substantial subset of the field has moved toward agents grounded in evidence-based psychological frameworks, particularly when the target population presents with comorbid mental health difficulties alongside their physical condition. Prochaska et al. (2021a, 2021b) deployed Woebot, a therapeutic relational agent adapted for substance use disorders, which drew on cognitive behavioural therapy, dialectical behaviour therapy, and motivational interviewing principles to support users in reducing problematic substance use. Hauser-Ulrich et al. (2020) evaluated SELMA, a chatbot delivering CBT-informed psychoeducation and pain self-management coaching for adults with chronic pain. Bruijnes et al. (2023) designed a conversational agent around personalised psychoeducation targeting social diabetes distress, the specific emotional burden arising from living with diabetes and the stigma it can carry. Across these examples, chatbots function not merely as information conduits but as behavioural support tools, structured to elicit reflection, reinforce coping skills, and provide a consistent, non-judgmental presence that clinical services often cannot offer at scale.

The interaction modalities through which these functions are delivered vary considerably and carry their own implications for engagement and experience. Text-based conversational interaction remains the dominant mode across the literature, delivered through smartphone applications, web platforms, and messaging services such as WhatsApp and Facebook Messenger (Cook et al., 2024; Ma et al., 2025). Voice-based interaction has been explored in more specific contexts, including with elderly adults with type 2 diabetes using an Amazon Echo device, where the conversational interface introduced a qualitatively different form of engagement compared with text-based alternatives (Ludwig da Costa et al., 2025). Embodied agents, such as the eEVA system evaluated by Boustani et al. (2021) for alcohol misuse, or the Laura agent used in diabetes self-management by Baptista et al. (2020), layer visual presence and nonverbal cues onto text or speech, enabling more nuanced communicative acts. Martinez-Miranda et al. (2019) illustrated this in a pilot study with adults with a history of suicidal ideation, in which an embodied agent capable of conveying emotional expressions through facial and gestural cues produced different user responses compared with a neutral presentation, underscoring the relevance of interaction modality to the user experience.

The breadth of chronic conditions addressed across this literature is itself noteworthy. Chatbots have been deployed for people managing type 2 diabetes, COPD, chronic heart failure, HIV infection, various cancers, asthma, chronic pain, chronic liver disease, autoimmune inflammatory diseases, and posttraumatic stress disorder, among others (Easton et al., 2019; ter Stal et al., 2021; Dworkin et al., 2019; Han et al., 2024; Hasei et al., 2025). This breadth reflects the structural logic that motivates the field: chronic conditions, by definition, unfold over time and require sustained engagement with self-management, monitoring, and emotional adjustment that no clinical system can provide continuously. Chatbots represent one plausible answer to this structural problem, precisely because they are available at any hour, require no appointment, impose no cost per interaction, and offer a degree of anonymity that some patients find genuinely enabling (Hasei et al., 2025; Wu et al., 2024). For populations such as young African American men living with HIV, for whom stigma and confidentiality concerns remain significant barriers to engagement with formal services, the private and non-judgmental character of a well-designed conversational agent may remove frictions that conventional care cannot (Dworkin et al., 2019).

What remains less settled, however, is what makes a conversational agent not merely usable but genuinely effective as a support tool in this context. The evidence suggests that technical capability and informational accuracy, while necessary, are not sufficient. Wu et al. (2024), in a regression analysis with adults managing chronic conditions, found that trust in a chatbot to provide appropriate support was the strongest attitudinal predictor of future adoption, surpassing both privacy concerns and comfort with symptom disclosure. Fewer than one third of their respondents indicated they were likely to use a health chatbot in the coming year, and their hesitancy was rooted not in doubts about functional performance but in uncertainty about whether a chatbot could respond to them in a way that felt sufficiently personal and trustworthy. Users across a wide range of studies consistently raised the quality of the relational interaction itself as a central determinant of their experience, their willingness to engage over time, and their sense of having been genuinely understood. The clinical impact of these tools, it appears, is inseparable from the degree of trust and interpersonal closeness that patients are able to develop with them. The issue, in other words, is not only whether these tools are accessible or clinically accurate, but whether they are perceived as sufficiently trustworthy, responsive, and empathic to support the kind of sustained engagement that chronic disease self-management requires. Understanding what those relational qualities consist of in chronic care contexts, and what conditions allow them to develop between a patient and a conversational system, is therefore the task the following sections take up.

1.3 Empathy, Trust, and Therapeutic Relationships in Chronic Care

Empathy occupies a foundational position in the theory and practice of healthcare communication. In clinical contexts, it shapes the depth of patient disclosure, the quality of the therapeutic encounter, and the degree to which patients feel understood, validated, and supported in managing their conditions. Sou et al. (2024), in developing a conceptual framework for empathy in healthcare settings, distinguish three interrelated subconstructs that together define what clinical empathy entails. Cognitive empathy is the capacity to understand another person's perspective and mental state without necessarily sharing their emotional experience, operating through a deliberate process of perspective-taking. Affective empathy is the more instinctive dimension, involving the direct sharing and resonance with another person's emotional state. Behavioral empathy is the communicative dimension through which these internal processes become perceptible to the patient: a clinician who experiences empathy but fails to express it through active listening, open-ended questioning, or appropriate verbal acknowledgment will not be perceived as empathic, and it is this behavioral dimension that most directly shapes the therapeutic encounter (Sou et al., 2024). Empathy in healthcare is therefore not simply an internal quality of the provider but a relational act that must be communicated to have clinical value.

This communicative dimension of empathy takes on particular significance in chronic disease management, where care unfolds across months and years of fluctuating symptoms, setback, and adaptation. Patients with long-term conditions must navigate their illness largely between clinical appointments, making decisions about self-management, coping with uncertainty, and managing the emotional weight of a condition that will not resolve. What patients require in this context is not only accurate information but a sustained sense of being understood and supported in a process that is inherently personal and emotionally demanding. Easton et al. (2019), in their co-design work with COPD patients, found that emotional support during low-mood periods was among the highest-priority self-management needs identified by participants, alongside symptom monitoring. The separation of physical and emotional care that characterizes much of chronic disease management does not reflect the lived experience of patients, for whom these dimensions are inseparable.

Central to the therapeutic encounter in chronic and long-term care is the concept of working alliance: the collaborative bond between patient and provider that encompasses shared therapeutic goals, agreed tasks, and an interpersonal attachment that sustains engagement and motivation over time. Working alliance is one of the most consistently replicated predictors of treatment outcome in psychological and behavioral health research, and its relevance extends beyond formal psychotherapy to any sustained helping relationship in which a patient must remain engaged, make behavioral changes, and tolerate uncertainty (Hauser-Ulrich et al., 2020). For patients with chronic conditions, the relational quality of their interactions with care providers is not peripheral to clinical outcomes; it is among the conditions that determine whether recommended behaviors are adopted, whether distress is disclosed, and whether the patient continues to engage with the care system at all. When that bond is strong, patients are more likely to follow guidance, voice concerns, and persist through difficulty. When it is absent or experienced as inadequate, the consequences extend to adherence, disease management, and health outcomes over the long term.

Trust operates alongside empathy as a structurally distinct but equally important relational construct in chronic care. Where empathy concerns the felt quality of understanding and emotional resonance in a clinical encounter, trust concerns the patient's confidence that the care provider is reliable, competent, and genuinely oriented toward their wellbeing. In chronic disease contexts, where patients interact with healthcare providers repeatedly over long periods and must rely on their guidance for complex and consequential decisions about their own health, trust is not a supplementary feature of the therapeutic relationship but a prerequisite for it. van Bussel et al. (2022), examining the determinants of healthcare tool acceptance among cancer patients, found that trust operated as an independent and significant factor in patients' willingness to engage with care, distinct from their assessment of performance or ease of use. Patients did not evaluate healthcare support purely on whether it worked; they also asked whether they could trust it, and the answer to that question shaped their willingness to engage at all. This finding speaks directly to the conditions of chronic care, where a patient who does not trust a source of support is unlikely to follow its guidance, disclose their difficulties, or return to it when they struggle.

The clinical importance of empathy and trust in chronic care is thrown into sharper relief by the scale of the gap between what patients need and what the current healthcare system provides. As the preceding sections established, the mental health burden among people with chronic conditions is substantial and the barriers to professional support are significant. The consequence is that many patients must manage the emotional dimension of their illness without adequate relational care, experiencing a form of support that is informationally adequate but relationally thin. MacNeill et al. (2024) made the depth of this unmet need concrete in their randomized controlled trial, in which adults with arthritis and diabetes who received structured mental health support over four weeks showed significant reductions in both depression and anxiety, a finding that points not only to the effectiveness of the intervention but to the extent of emotional need that had previously gone unaddressed. Lee et al. (2025), approaching the same gap from the patient's perspective, found that adults with anxiety consistently identified the absence of empathy as a primary concern when evaluating care options, articulating a clear and substantive expectation that effective support must feel genuinely responsive to their individual experience. The relational dimension of care is not, for these patients, a preference layered on top of clinical adequacy. It is a condition of acceptability.

It is against this background that the question of empathy in conversational AI systems has emerged as a significant area of inquiry. If empathy and trust are among the conditions that determine whether patients engage with care, disclose their distress, persist through difficulty, and ultimately benefit from support, then any digital tool that aspires to function as a meaningful care adjunct in chronic disease management must grapple seriously with these relational dimensions. The challenge is not only technical but conceptual: what empathy means when the interaction partner is not human, how it can be operationalized and communicated through text or voice, and how patients with chronic conditions perceive and respond to these qualities in practice. How the field has approached this challenge, what communication strategies and design features have been developed to convey empathy in chatbot interactions, and what the evidence reveals about their effectiveness, are the questions that the following section examines.

1.4 Digital Empathy in Chatbots: Concepts, Communication Strategies, and Design Features

Transposing the relational qualities that make human therapeutic encounters effective into automated conversational systems is not a straightforward technical problem. It is, first, a conceptual one. Empathy in human healthcare is a multidimensional, bidirectional relational act shaped by tone, timing, nonverbal attunement, and the accumulation of shared history between a patient and a provider. A chatbot cannot experience the affective resonance that underlies human empathy, and attempts to replicate this without theoretical clarity risk producing interactions that patients find incongruent or even unsettling. Sou et al. (2024) addressed this challenge directly by developing a conceptual framework that translates the three subconstructs of clinical empathy, namely cognitive, affective, and behavioral, into specific and testable verbal communication strategies that can be implemented in a text-based chatbot. Their central argument is that while the internal experience of empathy may not be replicable in a digital system, its behavioral expression can be operationalized through deliberate design choices, specifically through empathic self-awareness cues (ESA) and empathic active listening cues (EAL). ESA cues signal to the patient that the chatbot is responsive to its own communicative limitations and to the patient's emotional state, while EAL cues, including opening questions, probing, clarification, and evaluative reflections, create the conversational conditions under which a patient feels genuinely heard. This framework moves digital empathy from an abstract aspiration to a designable property, and its application in an experimental study with 921 adults with chronic conditions demonstrated that these specific verbal expressions positively shaped perceptions of chatbot quality and interaction appropriateness.

Beyond specific communication cues, the most consistently successful approach to empathic design in chatbots has been grounding the interaction in evidence-based psychological frameworks that were developed precisely to structure therapeutic conversations with warmth, responsiveness, and motivational sensitivity. Motivational interviewing, in particular, has served as a foundational design template across multiple chatbot interventions in chronic and behavioral health contexts. Boustani et al. (2021) built the eEVA embodied conversational agent around a motivational interviewing delivery structure, incorporating explicit MI principles of communicating empathy, supporting self-efficacy, and rolling with resistance rather than confronting ambivalence. Prochaska, Vogel, Chieng, Kendra, et al. (2021) adapted Woebot for substance use disorders, drawing on cognitive behavioural therapy, dialectical behaviour therapy, and motivational interviewing to structure empathic and therapeutically calibrated interactions. Dworkin et al. (2019) similarly grounded their relational embodied conversational agent for HIV-positive young men in motivational interviewing principles, designing the agent to be perceived by users as caring and understanding. Hauser-Ulrich et al. (2020) embedded CBT-informed psychoeducation into SELMA's conversational structure, creating a framework in which empathy was not an added layer but a property of the therapeutic process itself. What these approaches share is the insight that empathic communication does not arise spontaneously from chatbot interaction but must be deliberately structured through frameworks that specify how the agent responds to disclosure, validates difficulty, and supports the patient's sense of agency.

A distinct design trajectory within the field has pursued empathic communication through embodiment and multimodal interaction, extending beyond text to incorporate the visual and acoustic dimensions of human empathic expression. ECAs present as animated, human-like avatars capable of facial expressions, body gestures, and speech, enabling communicative acts that text-based systems cannot replicate. Boustani et al. (2021) documented this in their evaluation of eEVA, whose empathic style was implemented through a combination of spoken reflections, smiling facial expressions, forward lean, head nods, and hand gestures that were physiologically validated against the Facial Action Coding System. Among users with alcohol misuse, the large majority described the agent as warm and reported being comfortable disclosing their drinking behaviour to it, with most expressing greater comfort than they would have felt with their medical doctor. Baptista et al. (2020) found that the embodied Laura agent for diabetes self-management was widely described as empathic, friendly, and nonjudgmental, though a subset of users noted that incongruence between verbal and nonverbal behaviour reduced the authenticity of the interaction, pointing to the demands of coherent multimodal design. Martinez-Miranda et al. (2019) demonstrated experimentally that emotional expressions conveyed through an embodied agent were rated as significantly more competent than a neutral presentation by adults with a history of suicidal ideation. Voice-based interaction has extended these relational possibilities into the domestic sphere: Ludwig da Costa et al. (2025) found that elderly adults with type 2 diabetes using a voice-activated smart speaker for diabetes care and mental health support described the device in terms of humanization and emotional attachment, with some referring to it as a friend.

Across interaction modalities, the empirical evidence consistently identifies personalization as the mechanism through which chatbots sustain the relational qualities that make their support effective. The working alliance introduced in the preceding section, the collaborative bond of shared goals and interpersonal attachment that sustains therapeutic engagement, can be established between a user and a chatbot, but it appears to depend on the degree to which the system can adapt its responses to the individual patient's history, preferences, and current emotional state. Hauser-Ulrich et al. (2020) found that working alliance bond scores with the SELMA chatbot were significantly higher in the intervention group at follow-up and comparable to scores obtained in human-guided internet therapies, with participants specifically citing the empathic and reliable character of the interaction as a valued feature of their experience. Prochaska, Vogel, Chieng, Kendra, et al. (2021) similarly reported that affective bond scores with Woebot were substantially higher than agreement on specific tasks and goals, suggesting that users formed a genuine relational connection with the chatbot independent of its informational content. Inkster et al. (2018) found, in their real-world evaluation of the Wysa chatbot for depressive symptoms, that empathy-driven conversations were associated with higher levels of user engagement, while MacNeill et al. (2024) noted that users valued Wysa's kind personality and lack of judgment but that conversations perceived as generic or formulaic were experienced as invalidating. Mitchell et al. (2021) identified the fundamental tension structuring this domain: human health coaches were rated as clearly superior in empathic flexibility and in tailoring support through probing questions, while the chatbot performed better on consistency and in fostering patient autonomy. Personalization and relational responsiveness are therefore not merely desirable features of chatbot design but the conditions under which working alliance can develop, and their absence progressively erodes user engagement over time.

The landscape of digital empathy in chatbots has shifted substantially with the emergence of large language models and generative AI, which have expanded both the conversational capabilities available to developers and the expectations that users bring to their interactions with AI systems. Espinoza et al. (2025), in a comparative evaluation of two versions of the Ana chatbot for dementia caregivers in Peru, found that the generative AI-powered version (Ana-V2) scored significantly higher than the rule-based version (Ana-V1) on empathy, understanding, support, and dependability, providing direct empirical evidence that generative AI can meaningfully improve the perceived relational quality of chatbot interactions. Bai et al. (2025) introduced a more complex picture in an evaluation of GPT-4 in cancer consultations: a medical review panel rated the chatbot as demonstrating higher empathy than oncologists, while a patient review panel consistently favoured physicians, suggesting that clinicians and patients attend to different dimensions of empathic expression and that rater perspective must be explicitly considered when interpreting chatbot empathy findings. The broader social diffusion of generative AI capabilities has also introduced a new and largely unforeseen challenge for existing systems. Zhu et al. (2025) found that following the public release of ChatGPT, users of an established motivational interviewing chatbot rated it as significantly less empathic and less satisfying than users had before its release, demonstrating that patient expectations for conversational empathy are not stable but are recalibrated as generative AI capabilities become publicly visible. This finding has significant implications for how the field evaluates progress: what constitutes an adequate level of digital empathy is not a fixed standard but a moving target, shaped by the evolving reference points that users bring from their wider experience of AI interaction.

Together, these bodies of evidence paint a picture of a field that has made meaningful progress in conceptualizing and implementing empathic design in conversational agents for chronic disease management, but one that also faces unresolved challenges. The translation of human empathy frameworks into designable chatbot properties, the role of psychological grounding and multimodal embodiment, the centrality of personalization to working alliance formation, and the rapidly shifting baseline of user expectations all represent genuine advances in understanding. At the same time, the evidence reveals significant limitations: measurement inconsistencies that make comparisons across studies difficult, a temporal erosion of empathic perception when personalization is insufficient, and a persistent gap between the relational qualities patients most value and what current systems reliably deliver. The following section examines these challenges in greater depth, mapping the gaps in the current evidence base and identifying the questions that remain unanswered.

1.5 Challenges, Limitations, and Gaps in Empathetic Chatbot Research

One of the most persistent structural problems in the field is the inconsistency with which empathy is defined and measured across studies. Sou et al. (2024) identified this as a foundational challenge, describing the ambiguity surrounding empathy in healthcare chatbot research as a jingle-jangle problem in which overlapping and inconsistent terminology produces misaligned research outcomes that cannot meaningfully be compared. The practical consequence of this ambiguity is that studies operationalize empathy in strikingly different ways: some measure it through validated scales assessing therapeutic alliance or relational empathy, some through user satisfaction items that conflate empathy with general positivity, some through qualitative thematic analysis, and many do not formally measure it at all. Where validated instruments have been used, they capture different dimensions: the Bond subscale of the Working Alliance Inventory used by Hauser-Ulrich et al. (2020) assesses interpersonal attachment, the Average Consultation and Relational Empathy Measure used by Zhu et al. (2025) captures conversational empathy specifically, and the Feeling of Being Heard scale used by Bruijnes et al. (2023) targets a more situated dimension of perceived responsiveness. Without a shared measurement framework, claims about the presence or absence of digital empathy across studies cannot be aggregated with confidence, and the comparative evaluation of design strategies remains largely speculative.

A second challenge concerns the temporal erosion of empathic engagement when chatbot interactions fail to adapt to the patient's evolving clinical situation. User perceptions of relational quality are not static; they are subject to progressive decline when the system does not respond to changing circumstances, emotional states, or expressed preferences. ter Stal et al. (2021) documented this pattern with particular clarity in their four-month longitudinal study of the Sylvia embodied conversational agent with COPD and heart failure patients: while initial perceptions of the agent remained broadly neutral, participants' likelihood of following its advice declined significantly after three weeks of use, with qualitative feedback identifying non-personalized and contextually inappropriate responses as the primary driver. Cook et al. (2024), in their mixed-methods feasibility study of the Brisa asthma chatbot, found that limited conversational fluency reduced patients' perceptions of being genuinely understood. Han et al. (2024) identified personalization, along with privacy-sensitive interaction design and clearly defined role boundaries with formal clinical care, as critical design requirements articulated by individuals with PTSD, reflecting the understanding that relational credibility depends not only on initial design choices but on the system's capacity for sustained contextual responsiveness.

A third and arguably more fundamental limitation concerns what might be called the relational ceiling of current chatbot systems: the emotional and relational complexity beyond which automated conversational agents demonstrably cannot provide adequate support. Leung et al. (2025) encountered this ceiling directly in their mixed-methods study with metastatic breast cancer patients using the AIPL chatbot: patients found the system unable to offer hope, unable to address the relational complexities of managing their diagnosis within their personal relationships, and incapable of providing the emotionally sustaining engagement that their oncologists, despite their limited time, could offer. For these patients, the informational adequacy of the chatbot was not in question, but its emotional adequacy was, and the gap between the two was experienced as a significant limitation. van Bussel et al. (2022) documented a related concern among cancer patients evaluating a conceptual virtual assistant, who explicitly expressed worry that virtual tools lacked the empathy and compassion required for meaningful mental health support. Mitchell et al. (2021) framed this as a structural tradeoff rather than a design failure: human coaches were rated as clearly superior in empathic flexibility and in tailoring support through probing questions that the chatbot could not replicate, while the chatbot performed better on consistency and in fostering patient autonomy. The relational ceiling is therefore not simply a matter of current technological limitation but reflects a deeper question about what dimensions of therapeutic human relationship can be approximated in automated interaction and which cannot.

The integration of empathically oriented chatbots into clinical care also raises safety and implementation concerns that remain inadequately addressed. Bai et al. (2025) found that a clinically relevant proportion of GPT-4 responses in cancer consultations contained potentially adverse content attributable to misinterpretation of medical terminology or reliance on outdated clinical guidelines, a difference from human physicians that, while numerically small, carries significant weight in an oncology context where misinformation can have serious consequences. A related tension was identified by Coleman et al. (2025), whose randomized controlled trial found that a digital clinician avatar produced superior medication knowledge acquisition compared with standard nursing education, yet was simultaneously rated as less empathic and less trustworthy than the human nurse. This finding crystallizes one of the most important unresolved tensions in the field: informational efficacy and relational credibility do not automatically coexist in digital health tools, and improving a chatbot's capacity to deliver accurate clinical information does not necessarily enhance, and may in some contexts undermine, the relational qualities that drive patient trust and sustained engagement. Easton et al. (2019) identified a further safety concern in their co-design work with COPD patients: crisis support and emotional monitoring during low-mood periods were among the most prioritized functions by patients and health professionals, yet these are precisely the capabilities most difficult to implement safely in autonomous systems that lack the judgment to appropriately escalate or refer in complex clinical situations.

Longitudinal engagement with chatbot interventions presents a consistent practical challenge across the literature. He et al. (2025), in their study of the Roby smoking cessation chatbot, observed a progressive decline in engagement and satisfaction across sessions that could not be attributed to clinical outcomes alone, suggesting that novelty attrition may systematically limit the effectiveness of chatbot interventions that depend on sustained interaction over weeks or months. Retention figures across the corpus range from approximately 70% to 88%, but dropout is rarely analyzed in relation to empathy perception or relational quality specifically, leaving the mechanisms of disengagement poorly understood. Alongside engagement attrition, questions of sample representativeness and implementation equity constitute a significant gap in the current evidence base. Most included studies recruited predominantly female, white, higher-education populations, and findings from these samples may not generalize to the broader population of chronic disease patients who stand to benefit from digital health support. Wu et al. (2024) explicitly identified older adults and those with limited prior digital health experience as exhibiting substantially higher AI hesitancy, despite being among the most clinically relevant target groups for chatbot-delivered chronic disease support. Espinoza et al. (2025)

demonstrated that digital literacy, infrastructure access, and the absence of culturally sensitive and linguistically appropriate design constitute barriers that high-income country research has largely not engaged with, even as the global burden of chronic disease falls disproportionately on low- and middle-income settings where traditional care provision is most constrained.

A final consideration that the literature has only partially addressed is the role of implementation context in shaping how patients perceive and engage with chatbot empathy. The conditions under which a chatbot is introduced, the degree to which its use is endorsed by clinical providers, and the extent to which patients understand its role and limitations all appear to influence the relational quality of the interaction in ways that the chatbot's design alone cannot account for. Tan et al. (2023) provided empirical evidence of this in their survey study of autoimmune inflammatory disease patients, finding that patients' comfort with chatbot-based symptom screening increased significantly after meeting with a physician who endorsed its use, pointing to clinical affirmation as an underexplored but potentially powerful facilitator of chatbot acceptance. Han et al. (2024) reinforced this from the patient's perspective, noting that individuals with PTSD placed significant importance on clear boundaries between chatbot-delivered support and formal clinical care, and on reassurance that the chatbot would refer them to human providers when appropriate. Taken together, these findings suggest that the relational experience of chatbot interaction is not determined by design features alone but co-constructed by the implementation context, including the messaging of clinical providers, the framing of the tool's role, and the broader structure of care within which the interaction takes place.

The challenges reviewed in this section collectively point toward a field that is active and innovative but methodologically immature in its treatment of empathy as a clinical and relational construct. Most fundamentally, the literature remains fragmented in how empathy in chatbot interactions is defined, how it is measured, and how it is perceived by users living with chronic conditions. This fragmentation is not merely a methodological inconvenience; it is a structural obstacle to building cumulative knowledge about whether digital empathy works, for whom, and under what conditions. Measurement inconsistency prevents meaningful comparisons across studies. Personalization failures erode the relational engagement that empathic design is intended to create. The emotional ceiling of current systems remains largely uncharted. And the populations most in need of accessible chronic disease support remain the least represented in existing research. It is precisely this fragmentation, conceptual, methodological, and empirical, that the present scoping review is positioned to address.

1.6 Rationale and Objective of the Scoping Review

The evidence reviewed in the preceding sections reveals a field that has made genuine progress in deploying conversational agents for chronic disease management, yet one that remains fragmented in a fundamental respect: how empathy in these interactions is defined, how it is measured, and how it is perceived by users living with long-term conditions has not been systematically mapped. Studies vary enormously in how they conceptualize digital empathy, whether they measure it at all, what instruments they use when they do, and what they treat as evidence of an empathic interaction. The patient perspective, specifically how users with chronic conditions themselves perceive and respond to the empathic qualities of chatbot interactions, has rarely been placed at the center of this inquiry in a systematic way. This is the gap the present review is designed to address.

A scoping review is the methodologically appropriate response to this situation. Scoping reviews, as defined by the foundational framework of Arksey and O'Malley (2005) and subsequently refined by the Joanna Briggs Institute (Peters et al., 2020), are designed to map the breadth and character of evidence on a given topic, clarify key concepts and definitions, and identify gaps in an emerging or heterogeneous literature. The present review does not seek to establish the efficacy of a specific empathic chatbot intervention against a clinical comparator, a question better addressed by systematic reviews with narrow eligibility criteria and pooled effect sizes. Rather, its purpose is to map the range of existing evidence on how users with chronic conditions experience empathy in chatbot interactions, to identify what conceptual frameworks and design strategies have been applied, to document the factors that shape user perceptions, and to locate the inconsistencies and gaps that limit current knowledge. This mapping function is precisely what scoping reviews are designed for: they accommodate the methodological diversity of a young and rapidly evolving field, allow the inclusion of qualitative, quantitative, and mixed-methods studies, and produce an evidence map that can guide both future research design and the practical development of empathic chatbot systems for chronic disease management.

The review is organized around a principal research question: How do users with chronic conditions perceive and experience empathy in chatbot-assisted disease management, and what factors influence these perceptions? This question is operationalized through three sub-questions. The first asks what specific chatbot design features and technical strategies, including tone, language style, and personalization, have been employed to convey or simulate empathy in chronic disease management contexts. The second examines how perceived empathic interactions influence user outcomes, including trust, satisfaction, therapeutic alliance, and adherence to care plans. The third investigates which technological, contextual, and user-related factors, such as demographics, disease type, and interaction modality, shape the perception and experience of digital empathy.

The objective of this scoping review is therefore twofold. The first aim is descriptive: to provide a systematic and transparent account of the existing empirical literature on digital empathy in chatbot interactions for chronic disease management, mapping the populations studied, the conditions addressed, the design strategies employed, and the outcomes reported. The second aim is analytical: to identify recurring patterns, points of convergence and divergence, and the substantive gaps in current knowledge that must be addressed before digital empathy can be understood as a clinically meaningful and reliably deliverable dimension of chatbot-mediated care. In doing so, the review responds directly to what Wu et al. (2024) identified as one of the most pressing needs in the field: understanding the elements that foster human-AI trust and empathic connection is essential for ensuring that chatbots can fulfill their promise as support tools for the millions of people living with chronic conditions who currently lack adequate access to sustained, emotionally responsive care.

Chapter 2

Method

2.1 Study Design

This study was designed as a scoping review to systematically map and synthesize users' perceptions of empathy and their experiences during chatbot interactions in chronic disease management. This approach is effective for mapping a broad body of evidence, clarifying complex concepts such as "digital empathy," and identifying gaps in the research landscape. Unlike systematic reviews, a scoping review allows for the inclusion of diverse study designs, including qualitative, quantitative, and mixed-methods research, which is essential for capturing the multidimensional nature of user experience.

The conduct of this review adhered to the methodological framework proposed by Arksey and O'Malley (2005) and refined by the Joanna Briggs Institute (Peters et al., 2020). The process involved five stages: identifying the research question, identifying relevant studies, study selection, data charting, and collating, summarizing, and reporting the results. Additionally, the review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews (PRISMA-ScR) guidelines (Tricco et al., 2018).

2.2 Research Questions

The review aims to address the following principal research question: How do users with chronic conditions perceive and experience empathy in chatbot-assisted disease management, and what factors influence these perceptions? This question examines how chatbot features, user characteristics, and interaction contexts influence trust, satisfaction, and treatment adherence.

To structure the analysis, the principal question is operationalized through the following sub-questions:

1. What specific chatbot design features and technical strategies (such as tone, language style, and personalization) are employed to convey or simulate empathy in chronic disease management?
2. How do these perceived empathetic interactions influence user outcomes, including trust, satisfaction, therapeutic alliance, and adherence to care plans?
3. Which technological, contextual, and user-related factors (e.g., demographics, disease type, interaction modality) influence the perception and experience of digital empathy?

2.3 Eligibility Criteria

The eligibility criteria were defined using the Population, Concept, and Context (PCC) framework recommended by the Joanna Briggs Institute:

- **Population:** This review includes individuals across the lifespan managing chronic or long-term health conditions, as well as their caregivers. Conditions of interest include diabetes, cancer, HIV, asthma, COPD, heart failure, chronic pain, and inflammatory bowel disease. Studies involving pediatric, adolescent, and young adult populations were also considered.
- **Concept:** The focus is on perceived empathy and emotional connection during chatbot interactions. This includes strategies used by conversational agents to convey empathy, such as personalized language, communication cues, and the development of therapeutic alliance or affective bonds.
- **Context:** Eligible studies are situated within healthcare or self-management settings, including mHealth applications, telehealth services, and digital tools supporting chronic disease management and follow-up care.
- **Study Design:** Only peer-reviewed empirical studies employing qualitative, quantitative, or mixed methods designs and reporting primary data were included.
- **Publication Criteria:** Studies published in English from the year 2000 onward were included to reflect the development of digital health technologies and conversational agents.
- **Exclusion Criteria:** Studies were excluded if they focused solely on technical system architecture without evaluating user experience, involved only simulated users

or prototype testing, addressed acute or temporary conditions, or examined non-health-related chatbot applications. Non-primary literature, including reviews, editorials, and opinion papers, was also excluded.

2.4 Information Sources and Search Strategy

A literature search was conducted across five major academic databases: PubMed, Scopus, PsycINFO, CINAHL, and Web of Science. These databases provide a robust foundation for identifying research at the intersection of digital empathy and chronic disease management.

The search strategy was developed iteratively to capture the intersection of conversational agents, empathy and user experience, and chronic disease management. A combination of controlled vocabulary (such as MeSH terms in PubMed) and free text keywords was employed, using Boolean operators and truncation (such as chatbot, *empath*, chronic disease*) to maximize sensitivity. Search strategies were tailored to the specific syntax and indexing requirements of each database. The complete search strategies are provided in Appendix 1. An example of the search string is also provided to enhance transparency.

Example PubMed Search String:

```
((chatbot*[Title/Abstract] OR "conversational agent*"[Title/Abstract] OR "
    social bot*"[Title/Abstract] OR softbot*[Title/Abstract] OR "virtual
    agent*"[Title/Abstract] OR "automated agent*"[Title/Abstract] OR "
    automated bot*"[Title/Abstract] OR "virtual therap*"[Title/Abstract] OR
     "virtual assistant*"[Title/Abstract]) AND ("chronic disease*"[Title/
    Abstract] OR "chronic condition*"[Title/Abstract] OR "long-term
    condition*"[Title/Abstract] OR "disease management"[Title/Abstract] OR
    "healthcare"[Title/Abstract] OR "digital health"[Title/Abstract] OR "
    health management"[Title/Abstract] OR "chronic patient*"[Title/Abstract
    ] OR "chronic user*"[Title/Abstract] OR "health*"[Title/Abstract])) AND
     (empath*[Title/Abstract] OR "user experience*"[Title/Abstract] OR
    perception*[Title/Abstract] OR attitude*[Title/Abstract] OR "emotional
    intelligence"[Title/Abstract] OR "affective computing"[Title/Abstract]
    OR "relational agent*"[Title/Abstract] OR "emotional support"[Title/
    Abstract])
```

In accordance with JBI methodology and PRISMA-ScR guidelines, the final search was updated on August 5, 2025.

The search yielded the following number of records: PubMed (n = 543), Scopus (n = 81), Web of Science (n = 667), PsycINFO (n = 150), and CINAHL (n = 109). This structured approach enabled the inclusion of a broad range of study designs, enhancing the comprehensiveness of the evidence synthesis.

2.5 Study Selection

All identified records were exported to Zotero for reference management and the removal of duplicate entries, while Rayyan was employed for the initial title and abstract screening process. Following deduplication, the selection process was conducted in two sequential phases:

1. **Phase One:** Two reviewers independently screened the titles and abstracts of all retrieved records. Studies that did not meet the inclusion criteria, such as those focusing on healthy populations, acute conditions, or lacking a focus on digital empathy, were excluded.

2. **Phase Two:** The full texts of the remaining potentially relevant articles were retrieved and independently assessed for eligibility by two reviewers.

Any discrepancies between the reviewers were resolved through consensus-based discussion or by consulting a third reviewer. The entire selection process, including specific reasons for exclusion at the full-text stage, is documented and visually summarized in a PRISMA-ScR flow diagram.

2.6 Data Charting Process

A standardized data charting form was developed following the JBI methodology for scoping reviews. To ensure reliability, the form was pilot tested by two reviewers on a sample of five included studies. This preliminary phase allowed the research team to refine the extraction categories and ensure that all relevant data points related to digital empathy were accurately captured. Following the pilot test, the finalized version was used for the full extraction of all articles.

The extraction framework was organized into several key domains:

- **Study Details and Design:** Authors, publication year, country, methodology, sample size, and recruitment methods.

- **Population and Context:** Targeted chronic conditions, participant demographics, and healthcare settings.

- **Chatbot Characteristics:** Technical details including name, purpose, degree of automation, and mode of interaction.

- **Empathy Features and User Experience:** Conceptualization of empathy, design strategies, and measures used to assess emotional connection, trust, and therapeutic alliance.

- **Outcomes and Influencing Factors:** Impacts on satisfaction, engagement, and treatment adherence, along with technological or contextual influences.

- **Synthesis of Findings:** Theoretical frameworks, key findings related to digital empathy, and identified research gaps.

The extracted data were synthesized into a structured summary table to identify patterns across methodologically diverse studies.

2.7 Data Synthesis and Analysis

Data synthesis was conducted in accordance with the JBI methodological framework and PRISMA-ScR reporting guidelines. Extracted data were organized into predefined categories and transferred into a structured summary table. Descriptive statistics such as frequencies, percentages, and means were used to summarize the general characteristics of the included studies and sample sizes.

For the synthesis of qualitative data related to users' perceptions of empathy, a thematic analysis approach (Braun and Clarke 2006) was applied to identify recurring patterns and key themes. The entire process was integrated using a systematic narrative synthesis approach to ensure methodological transparency. This approach enabled a comprehensive mapping of emotional interaction strategies and their influence on user-centered outcomes, including trust, satisfaction, and long-term disease management.

2.8 Ethical Considerations

This study was conducted as a scoping review of secondary data from previously published academic literature (Bai et al., 2025; Yan et al., 2025). As the research did not involve direct interaction with human participants or the collection of original personal data, it was exempt from formal institutional ethical review due to its non-interventional nature.

To ensure ethical validity, the following measures were implemented:

- **Integrity of Primary Research:** Only studies demonstrating adherence to established ethical protocols and Institutional Review Board (IRB) approval were included (Coleman et al., 2025; Han et al., 2024; Martinez-Miranda et al., 2019).

- **Informed Consent:** The review confirmed that original researchers obtained written informed consent from all participants (Anmella et al., 2023; Coleman et al., 2025; Han et al., 2024).

- **Data Privacy:** All extracted data were handled in an anonymized format. Any potentially identifying information was redacted, and no personally identifiable information was recorded (Inkster et al., 2023; Rathnayaka et al., 2022; Yan et al., 2025).

- **AI-Specific Ethical Considerations:** The review evaluated whether studies implemented safety net protocols to detect crisis language or self-harm risks (Anmella et al., 2023; Hauser-Ulrich et al., 2020; Rathnayaka et al., 2022). Strategies to mitigate AI hallucinations and misinformation through human oversight were examined to ensure patient safety and medical accuracy (Coleman et al., 2025; Cook et al., 2024; Yan et al., 2025).

2.9 Conflict of Interest

The author declares that there are no conflicts of interest regarding the research, authorship, or publication of this study.

Chapter 3

Results

3.1 Study Selection

A methodical search was performed across PubMed, Scopus, Web of Science, PsycINFO, and CINAHL, initially yielding 1550 records. After the removal of duplicates, 887 studies remained for title and abstract screening. After applying predefined inclusion and exclusion criteria, 89 full-text articles were assessed for eligibility; however, 49 studies were excluded because the full text was unavailable. In the end, 40 studies met the criteria for inclusion and were added to the final synthesis. Figure 1 shows the PRISMA-ScR flow diagram for the study selection process, and Table 1 summarizes the most important characteristics and main findings.

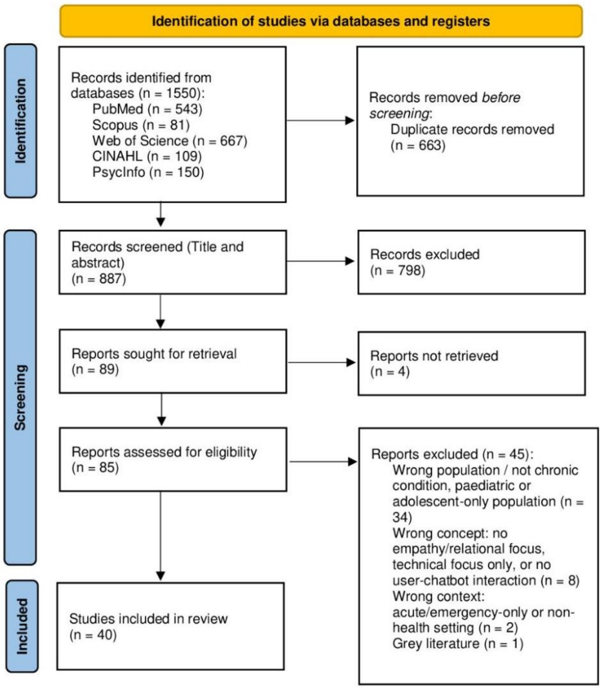

Figure 1: *Prisma flow-diagram of study selection and inclusion process*

3.2 Study Characteristics

The following table presents the key characteristics of the 40 studies included in the final synthesis, covering study design, population, chatbot details, and empathy-related findings.

Table 3.1: *Summary of Characteristics of Included Studies (N=40)*

Author(s), Year	Country	Study Method	Study Design	Sample Size (N)	Sample Characteristics	Population	Setting	Chatbot Name & Purpose	Interaction Mode	Empathy-Related Findings
Yan Z et al., 2025	China	Quantitative	Cross-sectional comparative study	263 questions; 1578 evaluations	NA	Patients with inflammatory bowel disease (Crohn disease and ulcerative colitis); evaluators included IBD patients and specialist doctors	Online patient education context using chatbot-generated and doctor-authored text	ChatGPT (GPT-3.5) used for patient education responses to IBD-related questions	Text-based question-and-answer	ChatGPT empathy scores were slightly lower than doctors before correction but showed no significant difference after Bonferroni correction
Sou D et al., 2024	Switzerland; Germany; United Kingdom	Quantitative	Online between-groups experimental study with 2x2 factorial design	921 patients	Mean age 42.4 years; 50.27% female	Adults with chronic conditions recruited from the UK	Online text-based healthcare chatbot in a simulated medical interview	Cary; prototype healthcare chatbot for conducting a medical interview	Text-based conversational interaction with predefined options and limited free text	NA

Author(s), Year	Country	Study Method	Study Design	Sample Size (N)	Sample Characteristics	Population	Setting	Chatbot Name & Purpose	Interaction Mode	Empathy-Related Findings
Mitchell EG; Maimone R; Cassells A; Tobin JN; Davidson P; Smaldone AM; Mamykina L, 2021	United States	Mixed-methods	Two-arm comparative field study with wizard-of-oz chatbot and human coach; interviews, questionnaires, and message log analysis	23 participants with type 2 diabetes; 4 human health coaches	Mean age 54.9 years; 75% female	Adults with type 2 diabetes from low-income, predominantly minority communities	Text-message–based virtual health coaching with supplemental mobile web app for self-tracking	t2.coach; scripted health coaching chatbot supporting goal setting and self-management for type 2 diabetes	Text-based conversational interaction (SMS), daily check-ins, goal-setting dialogs	Human coaches uniquely expressed empathy and expanded scope of support; chatbot lacked flexible empathic responses but still fostered supportive experiences
Rathnayaka P; Mills N; Burnett D; De Silva D; Alahakoon D; Gray R, 2022	Australia	Mixed-methods	Design and development study with participatory pilot evaluation (quasi-experimental pre–post and qualitative feedback)	318 users downloaded app; 34 participants met inclusion criteria for quantitative pilot evaluation	NA	Individuals with mental health issues, primarily depression and anxiety; working-age adults	mHealth smartphone application (Android and iOS)	Bunji; behavioural activation chatbot for mental health support and monitoring	Text-based conversational interaction with activity scheduling, mood check-ins, and surveys	Participants reported increased awareness, encouragement, and emotional resilience when using the chatbot

Author(s), Year	Country	Study Method	Study Design	Sample Size (N)	Sample Characteristics	Population	Setting	Chatbot Name & Purpose	Interaction Mode	Empathy-Related Findings
Wu PF; Summers C; Panesar A; Kaura A; Zhang L, 2024	United Kingdom	Quantitative	Cross-sectional web-based survey study	888 respondents	Median age 63 years; 64.6% female	Adults with chronic conditions, primarily type 2 diabetes; subset with long COVID; predominantly adults aged 55 years and older	Online health management context; hypothetical chatbot use scenarios	NA	Hypothetical text-based and voice-based chatbot interaction scenarios	Respondents with long COVID were significantly more interested in emotionally intelligent chatbots than those without long COVID
Cook D; Peters D; Moradbakhti L; Su T; Da Re M; Schuller BW; Quint J; Wong E; Calvo RA, 2024	United Kingdom	Mixed-methods	Mixed-methods feasibility study with pre–post evaluation over three iterative waves	150 adults with asthma	Mean age 39.36 years; 71% female	Adults diagnosed with asthma living in the UK	mHealth support delivered via WhatsApp and web-based chatbot	Brisa; asthma support chatbot for risk assessment, symptom tracking, education, and self-management support	Text-based conversational interaction with optional voice input for research purposes	Chatbot was perceived as helpful and supportive, but limited conversational fluency reduced perceived personalization

Author(s), Year	Country	Study Method	Study Design	Sample Size (N)	Sample Characteristics	Population	Setting	Chatbot Name & Purpose	Interaction Mode	Empathy-Related Findings
Tak YW; Park Y-E; Baek S; Lee JW; Chung S; Lee Y, 2024	South Korea	Mixed-methods	Secondary analysis of randomized controlled trial data with cross-sectional survey and open-ended responses	264 participants	Median age 54 years; 70% female	Adults with cancer (breast, colorectal, or lung cancer) undergoing or having completed treatment	Smartphone-based commercial health care applications	Multiple apps (Noom, WalkOn, Second Doctor, Efilcare) including chatbot or automated information features for health management	Smartphone app–based interaction with automated information and optional chatbot features	Positive or neutral attitudes toward apps were associated with greater perceived support and preference for chatbot-based interaction
van Bussel MJP; Odekerken–Schröder GJ; Ou C; Swart RR; Jacobs MJG, 2022	Netherlands	Mixed-methods	Sequential exploratory mixed-methods study (qualitative interviews followed by quantitative cross-sectional survey)	12 interview participants (8 former patients, 4 clinicians); 127 survey respondents	Interview patients mean age 58.5 years; survey respondents mean age 61.4 years; 54% female in survey	Adults diagnosed with cancer (various types) including former, current, and pre-treatment patients	Conceptual virtual assistant demonstrated via online mock-ups; online interviews and survey	Conceptual virtual assistant (no deployed name); intended for information provision, symptom self-diagnosis support, and mental health support for cancer patients	Text-based conversational virtual assistant (demonstrated via video mock-ups)	Participants expressed concerns that virtual assistants lack empathy and compassion, particularly for mental health support

Author(s), Year	Country	Study Method	Study Design	Sample Size (N)	Sample Characteristics	Population	Setting	Chatbot Name & Purpose	Interaction Mode	Empathy-Related Findings
Easton K; Potter S; Bec R; Bennion M; Christensen H; Grindell C; Mirheidari B; Weich S; de Witte L; Wolstenholme D; Hawley MS, 2019	United Kingdom	Mixed-methods	Participatory co-design study with qualitative workshops and quantitative usability assessment	6 participants with COPD and 5 health professionals in co-design workshops; 12 adults with COPD in acceptability testing	Median age 71–82.5 years across samples; mixed gender (exact percentages NA)	Adults with chronic obstructive pulmonary disease (COPD) and comorbid emotional difficulties; health professionals involved in COPD care	Home-based and community digital health context using a virtual agent	Avachat (Ava); autonomous virtual agent to support self-management, emotional support, crisis support, and motivation for COPD	Text-based and voice-based conversational interaction demonstrated via video scenarios (Wizard-of-Oz prototype)	Human-like presentation and emotionally supportive responses were viewed as acceptable and valuable; participants emphasized reassurance and feeling supported

Author(s), Year	Country	Study Method	Study Design	Sample Size (N)	Sample Characteristics	Population	Setting	Chatbot Name & Purpose	Interaction Mode	Empathy-Related Findings
Prochaska JJ; Vogel EA; Chieng A; Kendra M; Baiocchi M; Pajarito S; Robinson A, 2021	United States	Quantitative	Single-group pre–post feasibility and usability study	101 participants	Mean age 36.8 years; 75.2% female	Adults screening positive for problematic substance use (alcohol and/or drugs), many with comorbid depression or anxiety	mHealth smartphone application	Woebot (W-SUDs); therapeutic relational agent delivering CBT-based and motivational content for substance use reduction	Text-based conversational interaction with mood, craving, and pain tracking	Higher affective bond scores suggested users experienced the chatbot as relational and supportive
Han HJ; Mendu S; Jaworski BK; Owen JE; Abdullah S, 2024	united states	Qualitative	Semi-structured interview study with thematic analysis	12 participants	Age range 18–75 years; 9 female, 2 male, 1 non-binary	Adults living with post-traumatic stress disorder (PTSD)	Web-based conversational agent evaluated remotely in a home-based context	PTSDialogue; conversational agent to support PTSD self-management through assessment, psychoeducation, symptom management, and resource navigation	Text-based conversational interaction with predefined branching options and personas	Empathetic tone, personalization, and persona design enhanced acceptance and perceived usefulness of the chatbot

Author(s), Year	Country	Study Method	Study Design	Sample Size (N)	Sample Characteristics	Population	Setting	Chatbot Name & Purpose	Interaction Mode	Empathy-Related Findings
MacNeill AL; Doucet S; Luke A, 2024	Canada	Mixed-methods	Randomized controlled trial with embedded mixed-methods feedback analysis	68 participants	Mean age 42.87 years; 69% female	Adults diagnosed with arthritis or diabetes	mHealth smartphone application used in a home-based setting	Wysa; mental health chatbot providing automated psychological support and self-care exercises	Text-based conversational interaction with a combination of free-text and scripted responses	Users valued the chatbot's kind personality, lack of judgment, and reassuring presence, though some felt conversations were generic or invalidating
Inkster B; Sarda S; Subramanian V, 2018	United Kingdom; India	Mixed-methods	Real-world quasi-experimental pre–post evaluation with concurrent qualitative thematic analysis	129 users included in analysis (108 high users; 21 low users)	NA	Anonymous global users with self-reported depressive symptoms (PHQ-2 score of 6)	mHealth smartphone application used in real-world, in-the-wild context	Wysa; AI-driven mental well-being chatbot providing emotional support and evidence-based self-help techniques	Text-based conversational interaction with automated therapeutic tools	Empathy-driven conversations and tools were associated with higher user engagement and more favorable experiences

Author(s), Year	Country	Study Method	Study Design	Sample Size (N)	Sample Characteristics	Population	Setting	Chatbot Name & Purpose	Interaction Mode	Empathy-Related Findings
Anmella G; Sanabra M; Primé-Tous M; Segú X; Cavero M; Morilla I; Grande I; Ruiz V; Mas A; Martín-Villalba I; Caballo A; Esteva JP; Rodríguez-Rey A; Piazza F; Valdesoiro FJ; Rodriguez-Torrella C; Espinosa M; Virgili G; Sorroche C; Ruiz A; Solanes A; Radua J; Also MA; Sant E; Murgui S; Sans-Corrales M; Young AH; Vicens V; Blanch J; Caballeria E; López-Pelayo H; López C; Olivé V; Pujol L; Quesada	Spain	Mixed-methods	Development study with feasibility and potential effectiveness evaluation using pre–post self-assessments and usability questionnaires	40 healthy controls (setup); 17 healthy controls (simulation); 34 patients and health care workers (feasibility study)	Feasibility sample mean age 35.3 years; 76% female	Adults with anxiety-depressive symptoms and/or work-related burnout attending primary care or working as health care professionals	Smartphone-based mental health intervention in primary care and occupational health settings	Vickybot; chatbot for mental health screening, monitoring, psychological intervention delivery, and suicide risk detection	Text-based conversational chatbot with self-assessments, psychological modules, reminders, and alerts	NA

Author(s), Year	Country	Study Method	Study Design	Sample Size (N)	Sample Characteristics	Population	Setting	Chatbot Name & Purpose	Interaction Mode	Empathy-Related Findings
Hauser-Ulrich S; Künzli H; Meier-Peterhans D; Kowatsch T, 2020	Switzerland	Mixed-methods	Pilot randomized controlled trial with quantitative outcomes and qualitative content analysis	102 randomized participants (59 intervention; 43 control)	Mean age 43.77 years; 80.4% female	Adults with ongoing or cyclic chronic pain of at least 2 months' duration	mHealth smartphone application used in a home-based setting	SELMA (painSELf-MAnagement); text-based healthcare chatbot delivering CBT-based psychoeducation and pain self-management coaching	Text-based conversational interaction with predefined answer options and multimedia content	Working alliance bond scores were comparable to guided internet therapies; empathy/bond frequently cited as a positive aspect of the intervention
Inkster B; Kadaba M; Subramanian V, 2023	United Kingdom;United States	Mixed-methods	Real-world mixed-methods observational study with quasi-experimental pre–post comparison and thematic analysis	51 users for quantitative analysis; 10 users for qualitative thematic analysis	NA	Women who self-reported experiencing maternal events (pre-pregnancy, pregnancy, perinatal, or postpartum) with depressive symptoms	mHealth smartphone application used in a real-world, home-based context	Wysa; AI-enabled emotionally intelligent conversational agent for mental health and wellbeing support.	Text-based conversational interaction via mobile app	Higher engagement with the emotionally intelligent chatbot was associated with greater reductions in depressive symptoms and expressed emotional support themes

Author(s), Year	Country	Study Method	Study Design	Sample Size (N)	Sample Characteristics	Population	Setting	Chatbot Name & Purpose	Interaction Mode	Empathy-Related Findings
Gomaa S; Posey J; Bashir B; Basu Mallick A; Vanderklok E; Schnoll M; Zhan T; Wen K-Y, 2023	United states	Mixed-methods	Single-arm pilot mixed-methods feasibility and acceptability study	34 enrolled; 27 completed follow-up	Mean age 61 years; 56% female	Adults with gastrointestinal cancer undergoing intravenous chemotherapy	Home-based mHealth intervention using text messaging and chatbot interface	RT-CAMSS; chatbot-integrated system for symptom monitoring, self-care guidance, patient education, and emotional support during chemotherapy	Text messaging with integrated structured chatbot symptom checker	Encouraging and supportive messages helped patients cope emotionally and maintain positivity during treatment
Boustani M; Lunn S; Visser U; Lisetti C, 2021	United states	Mixed-methods	Descriptive feasibility, acceptability, and utility study with quantitative questionnaires and illustrative qualitative feedback	51 participants	Mean age 28 years; 32% female	Adults aged 21–55 years with heavy alcohol use (at least one binge drinking episode in past year), not currently in treatment	Web-based digital health intervention accessed remotely via internet-enabled devices	eEVA (empathic embodied virtual agent); to deliver brief motivational interviewing for alcohol misuse	Speech-enabled and text-based conversational interaction with embodied virtual agent (3D avatar)	Participants attributed multiple empathic and human-like qualities to the agent and reported positive relational experiences

Author(s), Year	Country	Study Method	Study Design	Sample Size (N)	Sample Characteristics	Population	Setting	Chatbot Name & Purpose	Interaction Mode	Empathy-Related Findings
Au J; Falloon C; Ravi A; Ha P; Le S, 2023	Australia	Mixed-methods	Prospective usability study with surveys and qualitative interviews	20 participants	Median age 55.5 years; 55% female	Adults with decompensated chronic liver disease (predominantly alcohol-related cirrhosis)	In-person usability testing using an iPad in hospital and outpatient clinic settings	Lucy LiverBot; AI chatbot to provide disease, medication, and nutrition-related education to improve health literacy in chronic liver disease	Text-based conversational chatbot with optional voice-to-text and emoji support	Conversational and companion-like qualities were viewed as key strengths, contributing to perceived emotional support and well-being
Chou YH; Lin C; Lee SH; Lee YF; Cheng LC, 2024	Taiwan	Quantitative	Single-arm pre–post development and usability study	35 participants	Mean age 65.21 years; 74% female	Older adults with diagnosed depressive or anxiety disorders receiving psychiatric outpatient care.	Home-based mHealth intervention delivered via LINE chatbot	Health Promotion chatbot; to provide companionship, monitor mood/sleep-/activity, and deliver health promotion education	Text-based chatbot interaction using button-based responses and daily prompts.	Older adults benefited from chatbot companionship, with significant improvement in UCLA Loneliness Scale scores in the 65 age group

Author(s), Year	Country	Study Method	Study Design	Sample Size (N)	Sample Characteristics	Population	Setting	Chatbot Name & Purpose	Interaction Mode	Empathy-Related Findings
Immel D; Hilpert B; Schwarz P; Hein A; Gebhard P; Barton S; Hurlemann R, 2025	Germany	Qualitative	Two-part qualitative survey study using web-based questionnaires and semistructured face-to-face interviews with grounded theory analysis	50 total participants (30 health care professionals; 20 patients)	Patients mean age 40.85 years; 45% female; health care professionals 80% female	Adults diagnosed with depression (patients) and health care professionals involved in outpatient psychiatric aftercare	Conceptual outpatient aftercare context; no deployed intervention (expectations-based study)	Virtual Therapeutic Agent (VTA); proposed socially interactive agent to support outpatient aftercare for depression and suicide prevention	Expected multimodal interaction including voice, facial expressions, gestures, and dialogue	Both patients and professionals emphasized emotional-supportive skills, particularly supportive-trusting and supportive-motivational behaviors, as central to effective VTAs
Pienkowska A; Ang CS; Mammadova M; Mahadzir MDA; Car J, 2023	Singapore	Mixed-methods	Participatory co-design study with iterative surveys and structured interviews	8 participants	Mean age 56 years; 38% female	Adults living with type 2 diabetes recruited from diabetes-related social media groups	mHealth app prototype evaluated remotely via web-based surveys and video interviews"	Diabetes education app prototype with rules-based chatbot; to support diabetes education and self-management skill	Text-based chatbot conversations with predefined options (no free-text input)	Opinions toward the chatbot were ambivalent and polarized, with some valuing perceived personalization and others criticizing lack of autonomy and free-text interaction

Author(s), Year	Country	Study Method	Study Design	Sample Size (N)	Sample Characteristics	Population	Setting	Chatbot Name & Purpose	Interaction Mode	Empathy-Related Findings
McAlister K; Baez L; Huberty J; Kerppola M, 2025	United states	Mixed-methods	Human-centered design study with qualitative ethnographic interviews and a quantitative pilot usability evaluation	43 participants in ethnographic interviews; 108 participants in pilot usability study	NA	Pregnant and postpartum individuals with mental health support needs during the perinatal period	Web-based chatbot accessed remotely in a home-based setting	Moment for Parents; rules-based chatbot providing perinatal mental health education, emotional support, coping strategies, and self-care guidance	Text-based conversational interaction with mood check-ins and tailored exercises	Human-centered design emphasizing validation and emotional support contributed to perceived relevance, engagement, and repeated re-engagement
ter Stal S; Sloots J; Ramlal A; op den Akker H; Lenferink A; Tabak M, 2021	Netherlands	Mixed-methods	Exploratory longitudinal real-life evaluation with questionnaires, log data analysis, and semistructured interviews	11 patients	Median age 70 years; 36% female	Adults with chronic obstructive pulmonary disease and chronic heart failure	Home-based eHealth self-management application	Sylvia; embodied conversational agent to support self-management, provide reminders, feedback, tips, and small talk	Text-based interaction with embodied conversational agent avatar	Empathic characteristics such as friendliness were valued, but lack of personalization and inappropriate feedback reduced perceived reliability and willingness to follow advice

Author(s), Year	Country	Study Method	Study Design	Sample Size (N)	Sample Characteristics	Population	Setting	Chatbot Name & Purpose	Interaction Mode	Empathy-Related Findings
Lee HS; Wright C; Ferranto J; Buttimer J; Palmer CE; Welchman A; Mazor KM; Fisher KA; Smelson D; O'Connor L; Fahey N; Soni A, 2025	United states	Qualitative	Semi-structured qualitative interview study with rapid qualitative analysis	29 participants	Age range 19–66 years; 72.4% female	Adults with self-reported mild to moderate anxiety receiving care within a US health system	Mental healthcare context using hypothetical and prototype AI conversational agents	AI conversational agents for mental health support (no specific deployed chatbot evaluated)	Text-based conversational interaction discussed conceptually and via brief prototype exposure	Perceived lack of empathy was a major barrier to acceptance; participants preferred AI as a supplement rather than a replacement for human therapists

Author(s), Year	Country	Study Method	Study Design	Sample Size (N)	Sample Characteristics	Population	Setting	Chatbot Name & Purpose	Interaction Mode	Empathy-Related Findings
Dworkin MS; Lee S; Chakraborty A; Monahan C; Hightow-Weidman L; Garofalo R; Qato DM; Liu L; Jimenez A, 2019	United states	Quantitative	Prospective single-group pre–post pilot study	43 participants enrolled; 32 completed follow-up	Median age 29 years; 0% female (all participants were men)	Young African American men who have sex with men living with HIV and receiving antiretroviral therapy	mHealth smartphone application used in a home-based context	My Personal Health Guide; embodied conversational agent to support HIV medication adherence, health literacy, motivation, and self-management skills	Embodied conversational agent with audio-visual interaction, touch-based navigation, and text/audio educational content	Participants perceived the embodied agent as caring and understanding, supporting acceptability and engagement

Author(s), Year	Country	Study Method	Study Design	Sample Size (N)	Sample Characteristics	Population	Setting	Chatbot Name & Purpose	Interaction Mode	Empathy-Related Findings
Baptista S; Wadley G; Bird D; Oldenburg B; Speight J; The My Diabetes Coach Research Group, 2020	Australia	Mixed-methods	Sequential mixed-methods acceptability study embedded within a randomized controlled trial	93 intervention participants; 66 survey respondents; 19 interview participants	Mean age 55–57 years; 47–50% female	Adults with type 2 diabetes living in Australia	mHealth smartphone app used in everyday home settings	Laura; embodied conversational agent delivering diabetes self-management education, feedback, and motivational support	Embodied conversational agent with voice and touch-based interaction via smartphone	Empathic, friendly, nonjudgmental support enhanced acceptability, though incongruence between verbal and nonverbal behavior reduced perceived authenticity for some users

Author(s), Year	Country	Study Method	Study Design	Sample Size (N)	Sample Characteristics	Population	Setting	Chatbot Name & Purpose	Interaction Mode	Empathy-Related Findings
Leung YW; So J; Sidhu A; Asokan V; Gancarz M; Gajjar VB; Patel A; Li JM; Kwok D; Nadler MB; Cuthbert D; Benard PL; Kumar V; Cheng T; Papadakos J; Papadakos T; Truong T; Lovas M; Wong J, 2025	Canada	Mixed-methods	Three-phase mixed-methods study with pre–post surveys and qualitative focus groups	42 participants consented; 36 completed post-survey; 10 participated in focus groups	Majority aged 40–64 years; 100% female	Adult women diagnosed with HR+/HER2 metastatic breast cancer	Home-based digital health intervention accessed remotely	Artificial Intelligence Patient Librarian (AIPL); chatbot providing conversational education, symptom guidance, and curated resource recommendations for metastatic breast cancer	Text-based conversational interaction with natural language input and resource recommendation	Participants felt AIPL could not adequately address emotional complexity, relational challenges, or foster hope, particularly for advanced disease

Author(s), Year	Country	Study Method	Study Design	Sample Size (N)	Sample Characteristics	Population	Setting	Chatbot Name & Purpose	Interaction Mode	Empathy-Related Findings
Martínez-Miranda J; Martínez A; Ramos R; Aguilar H; Jiménez L; Arias H; Rosales G; Valencia E, 2019	Mexico	Mixed-methods	Exploratory pilot acceptability study with quantitative questionnaires and qualitative open-ended feedback	18 enrolled; 12 completed the full pilot	Mean age 31.5 years; 61% female	Adults with a history of suicidal ideation, planning, or attempt but without recent suicidality or severe mental illness	Home-based mHealth smartphone application	HelPath embodied conversational agent; to monitor mood, support CBT-based activities, and detect/prevent suicide risk	Embodied conversational agent with text-to-speech, graphical interface selection, and non-verbal emotional expressions	Emotional competence ratings were significantly higher than neutral; empathy conveyed through consistent emotional behavior supported acceptability

Author(s), Year	Country	Study Method	Study Design	Sample Size (N)	Sample Characteristics	Population	Setting	Chatbot Name & Purpose	Interaction Mode	Empathy-Related Findings
Tan TC; Roslan NEB; Li JW; Zou X; Chen X; Ratnasari; Santosa A, 2023	Singapore	Quantitative	Cross-sectional survey study guided by the RE-AIM framework	200 participants	Mean age 48.44 years; 60.5% female	Adults diagnosed with or referred for suspected autoimmune inflammatory rheumatic diseases attending rheumatology outpatient clinics	Outpatient rheumatology clinic using a web-based chatbot	SingHealth RheumConnect; rule-based chatbot for symptom screening, triage, and patient education in rheumatology	Text-based chatbot interaction via internet-enabled devices	NA
da Costa FL; Matzenbacher LS; Maia IS; Gheno V; Brum MAB; de Barros LGB; Blank LM; Telo GH, 2025	Brazil	Qualitative	Qualitative interview study with inductive thematic content analysis and deductive coding	30 participants	Mean age 71.9 years; 56.7% female	Elderly adults aged 65 years with type 2 diabetes receiving care at a tertiary hospital	Home-based use of a voice-activated smart speaker	Amazon Echo Dot (Alexa); interactive virtual assistant programmed to support mental health promotion, diabetes self-care, reminders, and education	Voice-based conversational interaction	Strong sense of humanization and emotional attachment emerged, with some participants referring to the device as a friend or person

Author(s), Year	Country	Study Method	Study Design	Sample Size (N)	Sample Characteristics	Population	Setting	Chatbot Name & Purpose	Interaction Mode	Empathy-Related Findings
Ma Y; Achiche S; Tu G; Vicente S; Lessard D; Engler K; Lemire B; MARVIN Chatbots Patient Expert Committee; Laymouna M; de Pokomandy A; Cox J; Lebouché B, 2025	Canada	Mixed-methods	Uncontrolled single-group mixed-methods usability study with convergent design	28 participants completed quantitative component; 9 participated in qualitative interviews/-focus groups	Mean age 40.2 years; 17.9% female	Adults living with HIV, on antiretroviral therapy, receiving care at a Canadian tertiary HIV clinic	Telehealth/mHealth via Facebook Messenger used remotely	MARVIN (Minimal AntiRetroViral INterference); AI-based chatbot providing expert-validated information and reminders to support HIV self-management and ART adherence	Text-based conversational interaction via Facebook Messenger	Empathic conversational style and emotional safety facilitated positive attitudes toward use and behavioural intention to use

Author(s), Year	Country	Study Method	Study Design	Sample Size (N)	Sample Characteristics	Population	Setting	Chatbot Name & Purpose	Interaction Mode	Empathy-Related Findings
Coleman S; Lynch C; Worlikar H; Kelly E; Loveys K; Simpkin AJ; Walsh JC; Broadbent E; Finucane FM; O'Keeffe D, 2025	Ireland	Mixed-methods	Feasibility randomized controlled noninferiority trial with postintervention surveys and open-ended feedback	43 participants (27 intervention; 16 control)	Mean age 47.41 years; 79% female	Adults with overweight or obesity initiating semaglutide self-injection therapy	Onsite clinical education using a web-based digital clinician accessed via tablet device	Digital clinician (AI chatbot with avatar); to provide standardized education on semaglutide self-injection	Multimodal interaction combining text, voice, facial expressions, and question-and-answer dialogue via animated avatar	Despite higher knowledge acquisition, the digital clinician was perceived as less empathic and less trustworthy than human nurses

Author(s), Year	Country	Study Method	Study Design	Sample Size (N)	Sample Characteristics	Population	Setting	Chatbot Name & Purpose	Interaction Mode	Empathy-Related Findings
Prochaska JJ; Vogel EA; Chieng A; Baiocchi M; Maglalang DD; Pajarito S; Weingardt KR; Darcy A; Robinson A, 2021	United states	Quantitative	Two-arm randomized controlled trial with waitlist control	180 randomized participants (88 intervention; 92 waitlist control)	Mean age 40 years; 65% female	Adults screening positive for substance misuse (CAGE-AID >1) with alcohol and/or drug use problems	Home-based mHealth smartphone application	Woebot for Substance Use Disorders (W-SUDs); therapeutic relational agent delivering CBT-based tools to reduce substance misuse	Text-based conversational interaction via mobile app	Affective bond scores were significantly higher than agreement on tasks and goals, indicating strong perceived relational connection with the chatbot
He L; Basar E; Wiers RW; Antheunis ML; Krahmer E, 2025	Netherlands	Quantitative	Prospective single-arm longitudinal intervention study	102 participants	Mean age 24.7 years; 53% female	Adult smokers in the Netherlands motivated to quit smoking within two weeks	Web-based chatbot intervention accessed remotely via mobile or computer	Roby; chatbot delivering motivational interviewing and CBT-based smoking cessation support	Text-based conversational interaction	Therapeutic alliance remained moderate and stable but was not associated with abstinence outcomes

Author(s), Year	Country	Study Method	Study Design	Sample Size (N)	Sample Characteristics	Population	Setting	Chatbot Name & Purpose	Interaction Mode	Empathy-Related Findings
Bai X; Wang S; Zhao Y; Feng M; Ma W; Liu X, 2025	China	Quantitative	Comparative cross-sectional evaluation study using retrospective asynchronous text-based consultations	4257 asynchronous text-based consultation records	NA	Patients with cancer engaging in asynchronous text-based consultations; evaluators included oncologists and patients with cancer	Telemedicine using asynchronous text-based messaging	GPT-4 (OpenAI); to generate responses to patient education and medical decision-making queries from patients with cancer	Asynchronous text-based question-and-answer interaction	Medical review panel rated chatbot empathy higher than physicians in both scenarios, whereas patient panel consistently favored physicians
Zhu J; Dong A; Wang C; Veldhuizen S; Abdelwahab M; Brown A; Selby P; Rose J, 2025	Canada	Quantitative	Quasi-experimental comparative study using two independent cohorts (pre–post exposure comparison)	272 participants (143 pre-ChatGPT cohort; 129 post-ChatGPT cohort)	Mean age 29.22 years pre-ChatGPT and 32.75 years post-ChatGPT; approximately 50% female in both cohorts	Adult smokers (current or recent) recruited via an online research platform	Web-based chatbot used remotely on personal devices	MIBot; motivational interviewing chatbot designed to increase readiness to quit smoking	Text-based conversational interaction with scripted questions and LLM-generated reflections	Post-ChatGPT users rated MIBot significantly lower on CARE empathy scores; familiarity with ChatGPT did not positively predict empathy ratings

Author(s), Year	Country	Study Method	Study Design	Sample Size (N)	Sample Characteristics	Population	Setting	Chatbot Name & Purpose	Interaction Mode	Empathy-Related Findings
Bruijnes M; Kesteloo M; Brinkman W-P, 2023	Netherlands	Mixed-methods	Double-blinded between-subjects feasibility study with randomized control condition and longitudinal pre–post evaluation	156 participants (79 chatbot; 77 control)	Mean age 38.6 years; 47.4% female	Adults with self-reported diabetes (type I, type II, or other) recruited via an online crowd-working platform	Web-based conversational agent accessed remotely in a home-based context	Conversational agent support system; to provide personalized psychoeducation and coping strategies for social diabetes distress	Text-based conversational interaction over three weekly sessions	Chatbot users did not report significantly higher feeling of being heard compared with control, despite greater reduction in diabetes distress

Author(s), Year	Country	Study Method	Study Design	Sample Size (N)	Sample Characteristics	Population	Setting	Chatbot Name & Purpose	Interaction Mode	Empathy-Related Findings
Hasei J; Hanzawa M; Nagano A; Maeda N; Yoshida S; Endo M; Yokoyama N; Ochi M; Ishida H; Katayama H; Fujiwara T; Nakata E; Nakahara R; Kunisada T; Tsukahara H; Ozaki T, 2025	Japan	Mixed-methods	Pilot pre–post feasibility study with quantitative self-ratings and qualitative feedback	5 participants	Age range 13–early 20s; 60% female	Pediatric, adolescent, and young adult patients diagnosed with cancer (leukemia, osteosarcoma, Ewing sarcoma)	mHealth support via smartphone messaging application (LINE), hospital and home-based use	Two GPT-4–based generative AI chatbots (Kokoro and Aoi); to provide age-appropriate empathetic conversational psychological support	Text-based conversational interaction via smartphone messaging platform	Four of five participants reported reduced anxiety and stress; chatbot provided a psychologically safe space to disclose concerns

Author(s), Year	Country	Study Method	Study Design	Sample Size (N)	Sample Characteristics	Population	Setting	Chatbot Name & Purpose	Interaction Mode	Empathy-Related Findings
Espinoza F; Cook D; Da Re M; Cuba Fuentes MS; Butler CR; Calvo RA, 2025	Peru	Mixed-methods	Human-centered design study with stakeholder interviews, comparative prototype evaluation, satisfaction questionnaire, and field observations	7 caregivers; 2 dementia experts; additional field study with 28 community health workers and 36 older adults/relatives	Caregivers median age 57 years; 100% female caregivers	Informal caregivers of people with dementia in Peru; community health workers and older adults involved in field observations	Web-based chatbot accessed remotely; community and home-based caregiving context	Ana (Ana-V1 non-GAI; Ana-V2 GAI-powered); to provide dementia care information, behavioral symptom management guidance, and emotional support for caregivers	Text-based conversational interaction via web browser (prototype), with menu-based options and open-ended input in GAI version	Ana-V2 scored significantly higher than Ana-V1 on empathy, understanding, support, and dependability, indicating improved perceived empathy with generative AI

Given the extensive number of variables coded during the data extraction phase, the complete data extraction table is available upon request.

3.2.1 Geographic and Temporal Distribution of Studies

An analysis of the geographic distribution of the included studies (n=40) demonstrates a disproportionate concentration within high-income nations, comprising over 80% of the corpus. Conversely, emerging or middle-income regions remain underrepresented,although contributions from countries such as Brazil, Mexico, and Peru indicate an expanding global interest (Espinoza et al., 2025; Ludwig da Costa et al., 2025; Martinez-Miranda et al., 2019). Specific regional contributions highlighted in the primary literature include investigations conducted in Spain (Anmella et al., 2023), the Republic of Ireland (Coleman et al., 2025), and the United Kingdom (Wu et al., 2024). From a chronological perspective, the publication dates range exclusively from 2018 to 2025,with a significant surge in output observed since 2023, underscoring the nascent and rapidly evolving trajectory of this research domain.

3.2.2 Methodological Approaches

The research paradigms utilized across the reviewed literature (n=40) were systematically classified into four principal methodological designs. As illustrated in Figure 2, quantitative approaches and technical pilot studies formed the bulk of the methodologies, while qualitative and mixed-methods designs were utilized to a slightly lesser extent.

Figure 2: Bar chart of methodological designs

Within these broader classifications, specific methodologies varied significantly. For instance, Wu et al. (2024) deployed a cross-sectional, quantitative web-based survey, whereas Coleman et al. (2025) executed a rigorous randomized controlled noninferiority trial. Furthermore, investigations prioritizing feasibility and usability parameters are well-exemplified by the pilot study by Anmella et al. (2023) evaluating the "Vickybot" digital intervention.

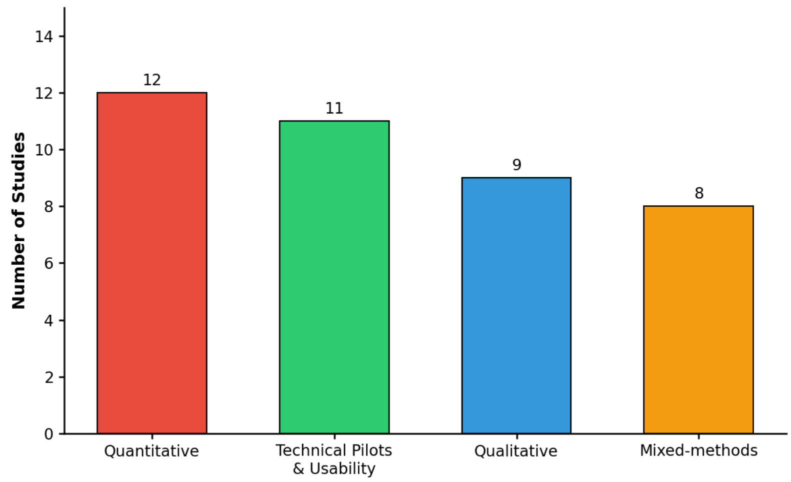

Figure 2: *Bar chart of methodological designs*

3.2.3 Clinical Domains and Target Conditions

The target health conditions addressed within the synthesized literature were stratified into four distinct clinical domains. Figure 3 highlights the distribution of these domains, indicating a primary focus on mental health and chronic condition management, with metabolic disorders and oncology representing smaller segments of the literature.

Figure 3: Horizontal bar chart of clinical domains

Within the predominant domains, Anmella et al. (2023) evaluated the efficacy of a chatbot intervention on anxiety, depressive symptoms, and burnout among healthcare professionals, noting potential efficacy in mitigating burnout. The chronic conditions category encompassed investigations into diverse long-term morbidities, including HIV, cirrhosis, and the management of Long COVID sequelae (Wu et al., 2024). Research in the metabolic sector focused primarily on patient education for anti-obesity pharmacotherapy (Coleman et al., 2025) and diabetes self-management protocols (Wu et al., 2024).

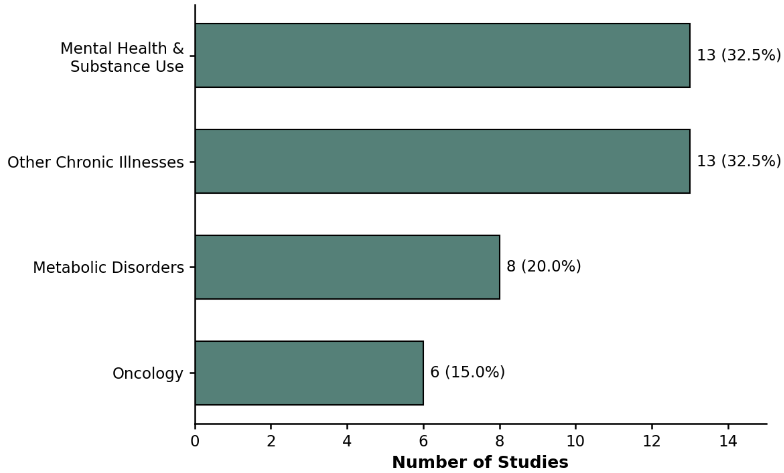

Figure 3: *Horizontal bar chart of clinical domains*

3.2.4 Demographic Characteristics of Participant Cohorts

Demographic synthesis of the included studies reveals a pronounced overrepresentation of female participants. This demographic skew is congruent with foundational literature, which estimates female engagement in digital health interventions to typically range between 70% and 86% (Hauser-Ulrich et al., 2020; Prochaska, Vogel, Chieng, Kendra, et al., 2021). Disaggregated data from the primary sources substantiate this trend across different geographical and clinical contexts (Figure 4).

Figure 4: Grouped bar chart showing gender distribution across key studies

Beyond gender, the targeted populations exhibited considerable heterogeneity. Wu et al. (2024) observed a cohort predominantly skewed toward older adults, with 75% of their sample aged 55 years and older. Participant cohorts also ranged from general primary care patients to specialized occupational groups, such as frontline healthcare workers (Anmella et al., 2023).

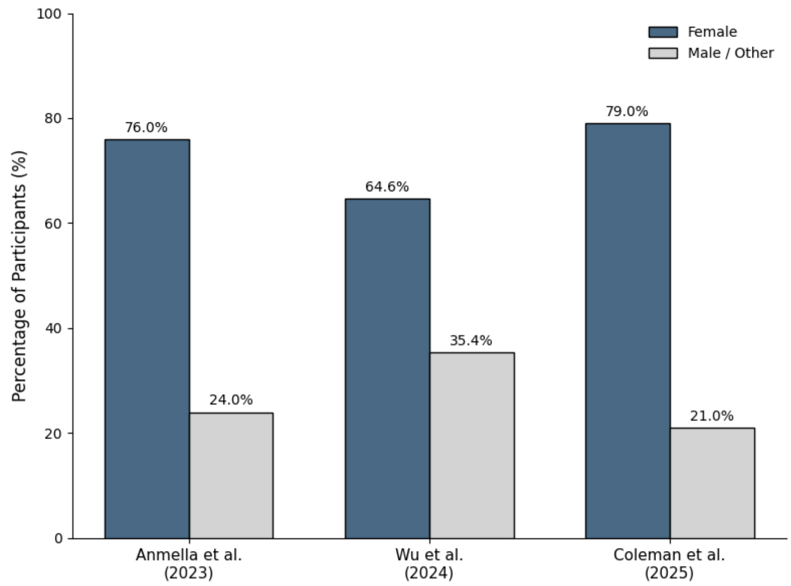

Figure 4: *Grouped bar chart showing gender distribution across key studies*

3.2.5 Theoretical Frameworks and Intervention Mechanics

The architectural design of the digital interventions analyzed relied upon a heterogeneous array of psychological theories and technological adoption frameworks. Therapeutically oriented applications, such as the one described by Anmella et al. (2023), integrated a composite of evidence-based psychological modalities, drawing upon Cognitive Behavioral Therapy (CBT), Mindfulness, Dialectical Behavior Therapy (DBT), and Acceptance and Commitment Therapy (ACT).

Conversely, pedagogically focused interventions prioritized clinical education and validated assessment metrics. For example, Coleman et al. (2025) utilized the Self-Injection Assessment Questionnaire (SIAQ), the Trust between People and Automation Scale (TPA), and the Patients' Overall Satisfaction with Primary Care Physicians Scale (CPSS) to evaluate intervention outcomes. From a human-computer interaction perspective, Wu et al. (2024) elucidated critical determinants of technology adoption, highlighting AI hesitancy, prior operational familiarity with conversational agents, and trust as primary mediators of user engagement.

Synthesized collectively, these findings illuminate a discernible paradigm shift within the digital health sector: a transition from exclusively therapeutic, symptom-management modalities toward comprehensive, education-centric modules that rigorously evaluate the psychosocial predictors of technology adoption and the cultivation of user trust.

3.3 Main Findings

The 40 studies included in this review address chatbots and conversational agents across a heterogeneous range of clinical contexts, most prominently chronic physical conditions such as type 2 diabetes, cancer, HIV, asthma, chronic pain, and COPD, alongside studies targeting mental health conditions and behavioural health issues including substance use and smoking cessation. Four recurring thematic domains emerge from the literature: clinical and behavioural outcomes, user experience and acceptability, empathy and relational aspects, and safety and implementation considerations. The following synthesis addresses each in turn, drawing out patterns of convergence and divergence and noting where the evidence remains limited or inconsistent.

3.3.1 Clinical and Behavioural Outcomes

Across studies reporting outcome data, empathic or relationally oriented chatbot interventions generally demonstrated favourable, though not uniformly robust, effects on clinically relevant endpoints. The strongest evidence came from controlled designs. MacNeill et al. (2024) conducted a randomised controlled trial in which adults with arthritis or diabetes used the Wysa chatbot over four weeks, finding significant reductions in depression ($p < .001$) and anxiety ($p < .001$) relative to a no-intervention control. Bruijnes et al. (2023) similarly demonstrated that a conversational agent delivering personalised psychoeducation on diabetes-related psychosocial distress produced a significantly greater reduction in Diabetes Distress Scale scores than a self-help book control ($p = .019$), with participants also scoring more favourably on a Feeling of Being Heard measure, suggesting a possible mediating role for perceived relational responsiveness.

Positive clinical signals were observed across other conditions, though with varying methodological rigour. Cook et al. (2024) reported an 8% improvement in self-reported asthma control over 28 days and high task completion rates among 150 participants using the Brisa chatbot. Dworkin et al. (2019) found antiretroviral adherence exceeding 80% and significant health literacy improvements among HIV-positive young adults using a relational embodied conversational agent and Gomaa et al. (2023) demonstrated significantly improved Patient Activation Measure scores in gastrointestinal cancer patients engaged with a chatbot-interfaced symptom management programme ($p = .02$), with 79% retention at two months. In the behavioural health domain, Prochaska et al. (2021a, 2021b) showed in both an uncontrolled pilot and a subsequent RCT (N=180) that the Woebot-SUDs adaptation significantly reduced past-month substance use occasions relative to waitlist control (p=.039).

Findings were less consistent elsewhere. Hauser-Ulrich et al. (2020) found that the

SELMA chronic pain chatbot did not produce a statistically significant improvement in pain-related impairment (p = .68) over eight weeks, despite participants reporting working alliance scores comparable to those in human-delivered internet therapies. This decoupling between relational engagement and clinical outcome is an important cautionary finding. He et al. (2025) observed that while the Roby smoking cessation chatbot achieved a continuous abstinence rate of 18.6% and significant self-efficacy improvements, engagement and satisfaction declined steadily across sessions, suggesting that novelty attrition may limit longer-term effectiveness. Coleman et al. (2025) reported that a digital clinician avatar significantly improved semaglutide self-injection knowledge in obesity patients ($p < .001$) compared with standard nursing education, yet trust levels were significantly lower in the digital clinician group ($p < .001$), highlighting a recurring tension between informational efficacy and relational credibility. Taken together, the available evidence supports the potential of empathic chatbots as adjuncts to chronic disease management, but effect sizes are often modest, follow-up periods short, and the absence of well-powered trials in most condition areas limits firm conclusions.

3.3.2 User Experience and Acceptability

User experience and acceptability constituted the most consistently positive domain across the corpus. Figure 5 illustrates representative acceptability and satisfaction scores from the studies reporting such data, revealing that most indicators clustered at or above 75%, with the notable exception of Wu et al. (2024), in which fewer than one third of surveyed adults with chronic conditions indicated they were likely to use a health chatbot in the next 12 months.

Figure 5. Overall user acceptability and satisfaction scores across included studies, expressed as a single representative percentage per study and ordered by condition category. The dashed line marks 50%; the dotted line marks 75%. Wu et al. (2024) is included as a contrasting benchmark reflecting general AI hesitancy in this population.

Beyond summary metrics, qualitative evidence revealed that users spontaneously formed relational constructs around chatbot interactions. Ludwig da Costa et al. (2025) reported that elderly individuals with type 2 diabetes described their voice assistant as having become "my best friend," with humanisation and companionship emerging as salient experiential themes. Baptista et al. (2020) found that 71% of adults with type 2 diabetes using the Laura embodied agent over 12 months rated it as helpful and friendly, and characterised it as a coach rather than a clinical tool. Au et al. (2023) noted that patients with decompensated cirrhosis perceived Lucy LiverBot as reliable and trustworthy, and reported improved emotional well-being as an unintended benefit of its conversational character.

Acceptability was not unconditional, however, and varied meaningfully over time. ter Stal et al. (2021) found in a four-month study of COPD and heart failure patients that while initial perceptions of the agent did not deteriorate significantly, participants were significantly less likely to follow its advice after three weeks (p = .01), attributing this to feedback they experienced as non-personalised. This finding demonstrates that surface-level acceptability can be maintained even as a chatbot's relational and behavioural influence erodes, a distinction that cross-sectional usability assessments are ill-equipped to detect.

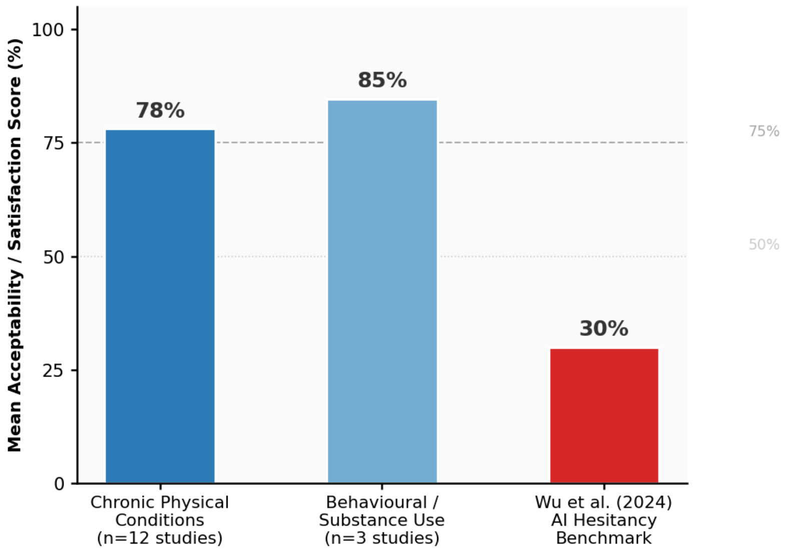

Figure 5: *Overall user acceptability and satisfaction scores across included studies, expressed as a single representative percentage per study and ordered by condition category. The dashed line marks 50%; the dotted line marks 75%. Wu et al. (2024) is included as a contrasting benchmark reflecting general AI hesitancy in this population.*

3.3.3 Empathy and Relational Aspects

Empathy and relational quality were the defining conceptual focus of this review, yet the included studies varied considerably in how these constructs were operationalised. Sou et al. (2024) provided one of the most theoretically grounded accounts in the corpus, developing a digital empathy framework for healthcare chatbots and testing it in an online experiment with an initial recruitment of 1,000 participants with chronic conditions (921 of whom provided complete data). Specific verbal empathic expressions, particularly empathic self-awareness and active listening behaviours, were found to positively shape perceptions of chatbot quality and interaction appropriateness, demonstrating that empathy can be deliberately designed and empirically verified rather than inferred post hoc from general satisfaction ratings.

Figure 6 summarises the proportion of empathy measurement approaches adopted across the 40 included studies, revealing a marked inconsistency in how the construct was operationalised.

Figure 6. Proportion of empathy measurement approaches across the 40 included studies. Nearly half of all studies (40%, n=16) did not formally measure empathy, treating it as incidental to other outcomes. User perception surveys were the most common explicit approach (25%, n=10), followed by qualitative methods alone (17.5%, n=7). Validated empathy or alliance scales were used in only 10% of studies (n=4), an expert or clinician rating panel in 5% (n=2), and an experimental manipulation of empathic conditions in a single study (2.5%).

The pattern in Figure 6 reveals a core challenge for the field: comparisons of empathy across studies are difficult to sustain when the construct is defined and measured so differently, or not measured at all. Where measurement did occur, user perception surveys and satisfaction items were the dominant approach, which, while pragmatic, do not distinguish between different dimensions of empathy such as cognitive understanding, emotional attunement, or therapeutic alliance. Validated scales, such as the Bond scale used by Hauser-Ulrich et al. (2020), the Average Consultation and Relational Empathy Measure used by Zhu et al. (2025), and the Feeling of Being Heard scale used by Bruijnes et al. (2023), appeared in only four studies, and an experimental design specifically manipulating empathic expressions was found in just one (Sou et al., 2024).

Despite these measurement limitations, several studies provided substantive findings on the relational character of chatbot interactions in chronic disease contexts. Mitchell et al. (2021) found that human coaches were rated as clearly superior in empathy expression and in tailoring support through probing questions, while the chatbot performed better on consistency and fostering patient autonomy, capturing a fundamental trade-off between empathic flexibility and relational predictability. van Bussel et al. (2022) demonstrated through structural equation modelling with cancer patients that trust was a significant predictor of virtual assistant acceptance independently of performance and effort expectancy ($\beta = 0.210, p < .001$), suggesting it functions as a distinct relational construct rather than a by-product of usability. Dworkin et al. (2019) found that relational design grounded in motivational interviewing principles and embodied avatar interaction improved health literacy among HIV-positive young adults, illustrating that theory-informed empathic design can translate into measurable behavioural gains.

Bai et al. (2025) introduced further complexity by demonstrating that the direction of the human-chatbot empathy comparison depended substantially on the rater: a medical review panel rated GPT-4 as demonstrating higher empathy than oncologists, while a patient review panel consistently favoured physicians. A similar ambiguity appeared in Yan et al.

(2025), where ChatGPT and specialist doctors received statistically comparable overall quality ratings in IBD care, but differences in empathy of borderline significance (p = .03) favoured specialist doctors, particularly in interactive consultation sections. These findings suggest that clinical evaluators and patients attend to different empathic dimensions, and that rater perspective must be explicitly considered when interpreting chatbot empathy findings.

Several studies also illuminated the temporal and contextual limits of chatbot empathy. ter Stal et al. (2021) found that non-personalised ECA feedback progressively undermined perceived reliability among COPD and heart failure patients. Leung et al. (2025) reported that metastatic breast cancer patients found the AIPL chatbot unable to offer hope or adequate relational support, pointing to an emotional ceiling that current systems do not reliably cross. Finally, Zhu et al. (2025) demonstrated that following the public release of ChatGPT, users rated an existing motivational interviewing chatbot as significantly less empathic (P=.01) and satisfying (P=.001) than pre-release users had, confirming that user expectations for conversational empathy are not fixed but are recalibrated as generative AI capabilities become publicly visible.

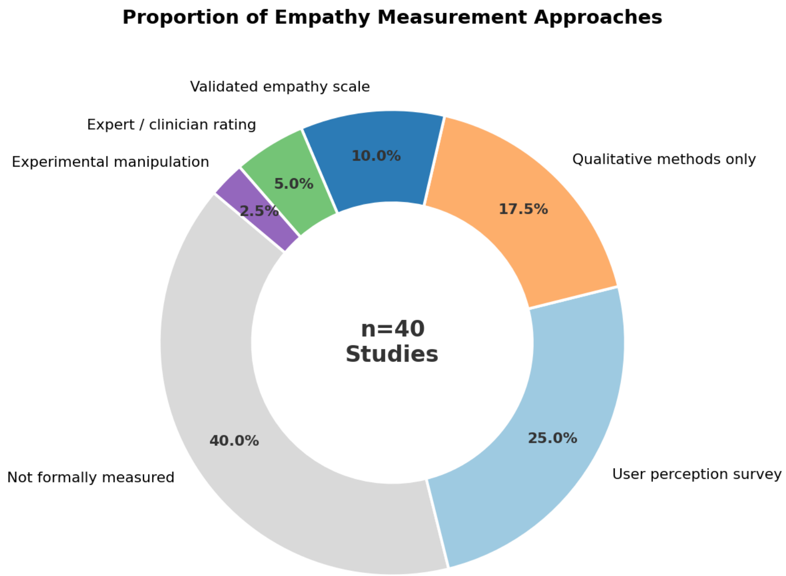

Figure 6: *Proportion of empathy measurement approaches across the 40 included studies. Nearly half of all studies (40%, n=16) did not formally measure empathy, treating it as incidental to other outcomes. User perception surveys were the most common explicit approach (25%, n=10), followed by qualitative methods alone (17.5%, n=7). Validated empathy or alliance scales were used in only 10% of studies (n=4), an expert or clinician rating panel in 5% (n=2), and an experimental manipulation of empathic conditions in a single study (2.5%).*

3.3.4 Safety, Limitations, and Implementation Considerations

Safety and implementation challenges were explicitly addressed in a subset of studies but represent an implicit concern across the corpus. Bai et al. (2025) reported that 3.12% of GPT-4 responses in cancer consultations contained potentially adverse content due to misinterpretation of medical terminology or outdated guidelines, compared with 2.01% for physicians, a small but clinically meaningful difference in an oncology context. Coleman et al. (2025) found that superior knowledge transfer from a digital clinician was accompanied by significantly lower trust, raising questions about appropriate informed consent framing when AI-generated health information is presented in clinical settings.

Personalisation emerged as the most pervasive implementation limitation. ter Stal et al. (2021) found that COPD and heart failure patients perceived ECA feedback as insufficiently individualised after three weeks, reducing their likelihood of acting on its advice. Han et al. (2024) identified personalisation, privacy-sensitive design, and clear role boundaries between the chatbot and formal clinical care as critical requirements articulated by individuals with PTSD, reflecting broader concerns about the appropriate scope of chatbot-delivered support. Easton et al. (2019) found through co-design with COPD patients that emotional support during low-mood periods and crisis response were prioritised functions, yet these represent precisely the capabilities most challenging to implement safely in autonomous systems.

Engagement attrition was a consistent practical challenge in longitudinal studies. He et al. (2025) observed declining engagement across the Roby smoking cessation intervention, while retention figures across the corpus ranged from 70% (Ma et al., 2025) to 84% - 88% (Prochaska, Vogel, Chieng, Baiocchi, et al., 2021), with drop-out rarely analysed in relation to empathy or relational quality, a gap in the current evidence base. Equity and generalisability concerns were also evident: most included samples were predominantly white, female, or higher-education populations. Wu et al. (2024) explicitly identified older adults and those with less digital health experience as exhibiting heightened AI hesitancy, despite being a core target group for chronic disease support. Espinoza et al. (2025) highlighted that digital literacy, infrastructure access, and culturally sensitive design remain substantial barriers that high-income country research has largely not engaged. Finally, Tan et al. (2023) observed that physician endorsement of chatbot use significantly increased patient comfort in autoimmune disease care ($p < .001$), pointing to clinical affirmation as an underexplored implementation facilitator.

In summary, the evidence collectively suggests that empathic chatbots hold genuine but context-dependent promise as adjuncts to chronic disease care, particularly in delivering accessible emotional and informational support and bridging gaps in care availability. The field is, however, characterised by significant methodological heterogeneity, small and often non-representative samples, short follow-up periods, and persistent inconsistency in how empathy and relational quality are operationalised. Rigorous longitudinal trials with standardised empathy measures, theory-grounded intervention design, and explicit attention to implementation equity are needed before the clinical value of empathic chatbots in this domain can be established with confidence.

Chapter 4

Discussion

The findings of this scoping review do not lend themselves to a simple verdict. Across 40 studies, users with chronic conditions seem to read chatbot empathy mainly through responsiveness, personalization, trust, and the feeling of being understood, and these readings shift depending on how the system is designed, what form the interaction takes, the clinical context in which it occurs, and how long the user has been engaging with it. The picture that emerges is one of real but conditional influence: chatbot design shapes how supported, trusted, and relationally engaged patients feel, yet the evidence so far stops short of showing that this translates reliably into lasting clinical improvement. This discussion interprets what the accumulated evidence actually means, connecting the patterns observed in the results back to the review's principal research questions and to the broader stakes of deploying emotionally responsive technology in the care of people with long-term health conditions.

4.1 Principal Findings in Relation to the Research Questions

The three sub-questions guiding this review asked, respectively, what empathic design strategies have been used in chatbots for chronic disease management, how perceived empathy shapes user outcomes, and which factors moderate those perceptions. It is worth reflecting on each in turn, not simply to restate the results but to consider what they collectively reveal about the current state of the field.

On the first question, the most striking observation is the gap between the sophistication of available empathy frameworks and how irregularly they appear in actual chatbot design. A small number of studies approached empathic design with genuine theoretical rigour, distinguishing between cognitive, affective, and behavioural dimensions and testing specific verbal cues empirically. The study by Sou et al. (2024) is the clearest example, but others incorporated motivational interviewing principles (Dworkin et al., 2019; Zhu et al., 2025), CBT-informed dialogue structures (MacNeill et al., 2024; Prochaska, Vogel, Chieng, Kendra, et al., 2021), or working alliance theory (Hauser-Ulrich et al., 2020) as explicit foundations. Many of the 40 studies, however, treated empathic tone as a design intuition rather than a theoretically grounded choice, something added to make the interaction feel warmer without clear operationalization of what warmth was supposed to achieve or how its presence would be verified. The result is a literature in which the word empathy describes a very wide range of things, from a single warm acknowledgment at the start of a session to a fully structured relational framework built around the therapeutic alliance.

The answer to the second question, about how perceived empathy influences outcomes, is where the evidence becomes genuinely interesting and also genuinely complicated. There is reasonable support for the view that empathic design improves engagement and relational quality: working alliance scores were meaningful and in some cases comparable to human-delivered therapies, users reported feeling supported and heard across a range of conditions, and in controlled designs there were significant improvements in depression, anxiety, adherence, and health literacy. However, the relationship between relational quality and clinical benefit is not straightforward. Hauser-Ulrich et al. (2020) found working alliance scores comparable to human therapy alongside a non-significant improvement in pain outcomes. He et al. (2025) found self-efficacy gains alongside steadily declining engagement. These dissociations matter because they suggest that the mechanisms connecting empathic interaction to health behavior change are more complex than the field has so far acknowledged, and they point toward an important limitation in the assumption that relational quality and clinical benefit will naturally travel together.

The third question, about moderating factors, may ultimately be the most consequential of the three. The results show clearly that perceived empathy is not a stable property of a chatbot but varies depending on how long a user has been interacting with the system, what disease they are managing, whether the system can adapt its responses over time, and, as became evident from the more recent studies, what other AI systems users have encountered before. These are not secondary considerations to be addressed in future iterations of the technology. They are the conditions under which empathic design either succeeds or fails to produce the relational quality it is intended to achieve, and understanding them more precisely should be among the field's priorities.

4.2 The Functionality of Empathic Design

There is sometimes a skepticism, not unreasonable, about whether a chatbot can be genuinely empathic in any meaningful sense, or whether what users experience is more accurately described as a simulation that sufficiently resembles empathy to produce similar effects. This review does not resolve that philosophical question, and perhaps it cannot. What it does suggest is that, for the practical purposes of chronic disease management, the distinction may matter less than it initially appears.

Sou et al. (2024) showed that specific, deliberate empathic expressions, structured active listening cues, expressions of self-awareness, and verbal acknowledgments of the patient's emotional state, produced measurable improvements in how users rated the quality of the interaction and the appropriateness of the chatbot's responses. Importantly, this was an experimental study with a manipulation check. It was not simply asking users whether they liked the chatbot; it was demonstrating that empathic cues, when added or removed, changed user perceptions in predictable ways. This reframes an important question in the field. Rather than asking whether chatbots can be empathic, the more tractable and practically relevant question is which communicative strategies reliably produce the experience of being understood.

The evidence reviewed here suggests that several such strategies are identifiable: acknowledging the emotional difficulty of living with a chronic condition rather than simply providing information, using the patient's own language and previous disclosures to frame new responses, asking questions before offering advice, and expressing something analogous to genuine concern when a patient reports a setback. Inkster et al. (2018), in their real-world evaluation of Wysa, found that interactions structured around these principles were associated with greater user engagement compared to more information-focused exchanges. MacNeill et al. (2024) found that users valued Wysa's consistent, non-judgmental tone, though they also found, pointedly, that when conversations felt scripted or generic, users experienced them as invalidating. The line between empathic and hollow is evidently not hard to cross, and what makes the difference is not the presence of warm-sounding language per se but whether the interaction feels genuinely responsive to the individual. What seems to hold across the evidence, then, is that empathic design is functional and consequential, but also fragile in ways that become more apparent with sustained use.

4.3 Personalization and Working Alliance: The Central Mechanism

If there is a single finding from this review that carries the most practical weight, it is probably this: personalization is not merely a technical feature added to make a chatbot feel friendlier, but a relational condition for whether engagement can survive past the first few weeks of use. Chronic conditions do not stay still. A patient's symptoms, mood, and daily circumstances shift over months and years, and a chatbot that cannot register and respond to that shift will eventually feel less attuned to the user than it did at the start, no matter how carefully its initial design was built. The relational quality of chatbot interactions, then, has less to do with which empathic strategies are present at the outset than with whether the system can sustain and adapt those strategies over time in response to the individual patient. This point emerges with notable consistency across studies that differed substantially in their methods, populations, and designs.

Hauser-Ulrich et al. (2020) found that working alliance bond scores with the SELMA chatbot were significantly higher in the intervention group and reached levels comparable to human-delivered internet therapies, a finding that sits somewhat against the conventional assumption that genuine therapeutic alliance requires a human interlocutor. Prochaska, Vogel, Chieng, Kendra, et al. (2021) found that affective bond scores with Woebot substantially exceeded agreement on specific tasks or goals, meaning that users were forming something recognizable as a relational connection with the system beyond its informational function. What these systems had in common was a consistent, theoretically grounded relational character that gave users a predictable and coherent interactional experience over time. That coherence, it seems, is part of what makes the alliance possible.

ter Stal et al. (2021) showed what happens when that coherence is not accompanied by genuine responsiveness to change, and the case makes the earlier point concrete. In their four-month longitudinal study with COPD and heart failure patients, users' willingness to follow the Sylvia agent's advice dropped significantly after three weeks. The agent had not changed; the users had. Their condition had evolved, their circumstances had shifted, and the system had no way of recognizing or responding to this. Participants did not feel that the agent knew them, and over time that feeling eroded their trust in its guidance. This is arguably one of the more important findings in the corpus because it explains why high initial acceptability ratings may mislead researchers about the long-term viability of a chatbot intervention. A system that feels helpful at week one may feel indifferent at week four, and the difference has nothing to do with the technology changing and everything to do with the user changing in ways the technology cannot track.

van Bussel et al. (2022) added a structural dimension to this picture through their work with cancer patients. Using structural equation modelling, they found that trust was a significant predictor of virtual assistant acceptance, with a standardized coefficient of 0.210, independently of perceived performance and effort expectancy. This is not a marginal result. It means that trust in this context is functioning as something more than a product of usability; it is a distinct relational judgment that users are making about whether the system can be relied upon as a genuine partner in care. Trust of that kind appears to be built, and potentially lost, through the cumulative quality of sustained interaction rather than through initial design choices alone.

4.4 Human and Chatbot Support: Towards a Complementary Model

The question of how chatbots compare to human care providers in empathic terms comes up repeatedly in the included studies, and the answers are more nuanced than either optimistic or pessimistic positions tend to acknowledge. Mitchell et al. (2021) provided perhaps the most direct comparison: human health coaches were rated clearly superior in empathic flexibility, in the ability to tailor responses through probing questions, and in the depth of understanding they conveyed. The chatbot, by contrast, was rated higher on consistency, on adherence to the therapeutic structure, and on fostering patient autonomy through systematic choice offering. Neither modality was straightforwardly better, and the differences between them suggest a division of strengths rather than a simple hierarchy.

Bai et al. (2025) introduced a further complication when they asked a medical review panel and a patient review panel to evaluate the empathy of GPT-4 responses versus oncologist responses in cancer consultations. The two panels disagreed: clinicians rated GPT-4 as more empathic, while patients consistently preferred their physicians. Yan et al. (2025) observed a similar pattern, though less pronounced, in ratings of ChatGPT versus specialist doctors in IBD consultations. Taken together, these findings suggest that empathy is not a unitary quality that all observers agree on recognizing. Clinicians and patients appear to attend to different dimensions of empathic expression, and a system that satisfies a clinical evaluator's criteria may still fall short of what a patient actually needs from an empathic interaction. This distinction has direct implications for how chatbot empathy should be evaluated and by whom.

Leung et al. (2025) put the limits of chatbot empathy in particularly sharp relief through their mixed-methods study with metastatic breast cancer patients. The AIPL chatbot was informationally adequate by most measures, and patients could obtain answers to factual questions about their condition and treatment. Emotional adequacy was another matter. Patients found the system unable to offer hope, unable to engage with the relational complexity of managing a life-altering diagnosis within a family, and incapable of the sustained attentiveness that their oncologists, however time-pressed, could still provide. Findings of this kind do not support treating chatbots as a substitute for human clinical relationships, and this should be stated explicitly rather than left as an implication. In emotionally complex or high-stakes contexts, where a diagnosis affects a patient's sense of the future or the wellbeing of their family, the evidence reviewed here offers no indication that an automated system can take the place of a clinician or a human carer. What it supports instead is a model of complementarity, in which chatbots address specific gaps in care rather than substitute for human contact more broadly.

That framing is supported by the outcomes literature, although the same literature also clarifies the conditions under which complementarity is achieved rather than simply assumed. Dworkin et al. (2019) demonstrated that a relationally designed embodied conversational agent grounded in motivational interviewing could improve antiretroviral adherence and health literacy in HIV-positive young adults in a context where access to frequent human therapeutic support was limited by cost and availability. Prochaska, Vogel, Chieng, Baiocchi, et al. (2021), similarly, found meaningful reductions in substance use occasions with Woebot-SUDs in a population that would otherwise have had limited access to structured behavioral support. What these studies share is a deployment context in which the chatbot was filling a gap that human care could not cover, rather than competing with support that was already available. The adjunctive role chatbots can play in such circumstances is genuine, but it depends on conditions that are rarely made explicit in the included studies: clearly defined boundaries on what the system is expected to handle, an operative pathway for escalating a patient to human care when the situation requires it, and continued clinical oversight rather than the delegation of emotionally complex cases to an automated system. In the absence of these conditions, the distinction between a chatbot that extends limited human support and one that effectively replaces it becomes considerably harder to maintain.

4.5 The Impact of Generative AI on User Expectations

Several of the more recent studies in this corpus capture what amounts to a natural experiment in how rapidly evolving AI capabilities reshape user expectations, and the results carry implications that extend well beyond the individual studies concerned. Zhu et al. (2025) studied users of MIBot, a structured motivational interviewing chatbot for smoking cessation, across two cohorts: one recruited before the public release of ChatGPT and one recruited after. The post-ChatGPT cohort rated MIBot as significantly less empathic and significantly less satisfying, despite interacting with exactly the same system. Familiarity with ChatGPT was directly and positively associated with lower empathy ratings. The interpretation is not complicated: users who had experienced a more capable and fluent generative system were applying a different standard when evaluating a structured, scripted chatbot, and that standard was higher.

Cook et al. (2024) observed something similar during the iterative development of the Brisa asthma chatbot across three testing waves in 2023. The first wave was conducted before generative AI had become a mainstream public experience; subsequent waves were conducted in a world where a substantial proportion of participants had used generative AI tools in their everyday lives. The drop in user satisfaction across waves was attributed in part by the research team to this contextual shift. What constituted an adequate chatbot conversation at the start of the study felt increasingly limited and rudimentary by its end, not because the system had deteriorated but because users' reference points had moved.

This dynamic has real implications for how the field interprets its own progress. A chatbot that achieves satisfactory empathy ratings today is not necessarily achieving them against the same standard that will apply in two or three years. The benchmark is moving, and it is moving because the broader population is gaining firsthand experience of what advanced conversational AI can do. Systems developed and evaluated before this shift may look considerably less impressive to future users than their original evaluations suggested, which in turn raises questions about the generalizability and durability of any empathy findings tied to a specific technological moment.

The case for generative AI as a solution to the empathy gap is not, however, without complications of its own, and these complications go beyond the technical errors that generative systems are known to produce. Espinoza et al. (2025) found that the generative AI-powered version of the Ana chatbot for dementia caregivers in Peru scored significantly higher than its rule-based predecessor on empathy, understanding, support, and dependability, which provides direct empirical support for the view that generative architectures can meaningfully improve the relational quality of chatbot interactions. At the same time, Cook et al. (2024) noted candidly that large language models continue to hallucinate information, produce confident errors, and generate false references in ways that patients cannot easily detect. There is a further risk that follows directly from the same fluency that makes these systems feel more empathic. The more naturally a chatbot communicates, the more it may invite a level of trust that its actual reliability does not justify, and a patient is not always able to distinguish a confident answer from an accurate one. The field's central challenge in the near term is to keep the relational gains that generative AI seems to offer without quietly inheriting these risks alongside them.

4.6 Implications for Chatbot Design and Clinical Practice

It would be premature to distill this review into a definitive set of design rules, given both the methodological heterogeneity of the evidence and the pace at which the technology is changing. That said, several findings are consistent enough across the corpus, and theoretically coherent enough, to carry genuine weight, and they are perhaps easiest to follow when separated according to who they speak to: those designing these systems, those deploying them clinically, and those setting the field's research agenda going forward.

4.6.1 Implications for Chatbot Design

The most robust implication concerns the case for theory-grounded design. Systems anchored in established clinical frameworks consistently demonstrated stronger relational and clinical outcomes than those developed without explicit theoretical grounding. Motivational interviewing shaped the relational architecture of Dworkin et al.'s (2019) embodied agent and of the MIBot studied by Zhu et al. (2025). Working alliance theory underpinned the SELMA intervention evaluated by Hauser-Ulrich et al. (2020). Cognitive behavioral frameworks structured the interactions in MacNeill et al.'s (2024) Wysa trial. Therapeutic theory encodes accumulated clinical knowledge about what kinds of communicative behavior actually help people, and translating that knowledge into chatbot interaction design requires effort and expertise but appears, on the evidence available, to be worth it.

Closely related is the case for personalization as a design priority rather than an optional enhancement. What repeatedly distinguishes systems that sustain user engagement over time from those that do not is whether the system can recognize and respond to the individual user's changing situation. This requires more than personalizing the greeting or remembering the user's name. It requires the system to track the trajectory of the interaction, notice when something has changed, and adapt accordingly. Han et al. (2024) found that individuals with PTSD articulated exactly this requirement when asked what would make a mental health chatbot trustworthy: personalization, contextual sensitivity, and a clear sense that the system understood the boundaries of its own competence. These are not aspirational features; they are conditions of relational credibility.

4.6.2 Implications for Clinical Practice

For clinical practice, the implications are perhaps more about realistic expectation-setting than about specific deployment recommendations. The evidence supports the use of empathic chatbots as adjuncts to chronic disease care, particularly for conditions where access to frequent human support is limited by cost, availability, or geography, and this review identifies several conditions under which that adjunctive role can be genuinely beneficial. It does not, however, support their use as substitutes for human clinical relationships in high-stakes contexts where those relationships are available and where the severity of the condition warrants sustained emotional engagement. Acknowledging that distinction clearly, both in system design and in clinical deployment decisions, seems more honest and ultimately more useful than designing around it.

4.6.3 Implications for Future Research

Finally, and perhaps most importantly for the field's own cumulative development, the results of this review point strongly to the need for shared measurement standards. Nearly half of the included studies did not formally measure empathy at all, and those that did used instruments that capture different constructs, some focused on interpersonal attachment, some on conversational responsiveness, some on the situated feeling of being heard. Without a common measurement framework, the evidence base will continue to grow in ways that are difficult to synthesize and impossible to aggregate quantitatively. Developing instruments specifically adapted for chatbot empathy in health contexts, and establishing minimum measurement conventions for studies in this domain, would be among the most practically impactful steps the research community could take.

4.7 Strengths of the Review

Before turning to these constraints, it is worth pausing on what this review was able to accomplish. The literature on chatbot empathy in chronic disease management has grown quickly and unevenly over the past several years, spread across disciplines that rarely cite one another and journals that rarely overlap in readership. Bringing forty studies from this scattered landscape into a single, structured synthesis is itself a contribution, since patterns such as the recurring tension between personalization and scalability, or the shifting baseline created by generative AI, only become visible once the evidence is assembled side by side.

There is also a strength in where this review chose to place its attention. Much of the existing literature on chatbot empathy approaches the question from the side of system design or clinical outcome measurement, treating the user's subjective sense of being understood as secondary to whether the intervention worked. This review kept that subjective experience at the centre instead, asking not only whether chatbots produced measurable health benefits but how patients themselves perceived and responded to their empathic qualities. Perceived empathy is not a soft adjunct to clinical effectiveness here. It is part of what determines whether patients keep using the system at all.

Finally, the review does not confine chatbot empathy to a single domain. The included studies span technical design choices, relational constructs such as working alliance and trust, and concrete clinical applications across a wide range of chronic conditions. Holding these dimensions together is what allows the preceding discussion to speak to developers, clinicians, and researchers at once.

Chapter 5

Limitations

The preceding discussion has drawn on the findings of this scoping review to interpret what the evidence means for empathic chatbot design and for the care of people with chronic conditions. It would be incomplete, however, without an honest account of the constraints that limit what can be concluded from this review, both in terms of its own methodology and in terms of the nature of the evidence it was able to include. Several of these constraints are significant enough to shape how the findings should be read.

5.1 Limitations of the Review Methodology

As with all scoping reviews, this study maps the landscape of available evidence rather than evaluating its quality in a formal sense. Unlike systematic reviews with meta-analytic components, scoping reviews do not apply standardized quality appraisal tools to individual studies, and the present review is no exception. This means that stronger and weaker studies have been synthesized alongside one another, and the narrative conclusions drawn here reflect the overall character of the literature rather than a weighted assessment of its methodological rigour. Some of the findings described, particularly those from feasibility studies and single-arm pilots without control conditions, carry considerably less evidential weight than others, and readers should bear this in mind when interpreting the review's conclusions.

The search strategy, while systematically designed and applied across five major databases, was restricted to English-language publications. This decision was made for pragmatic reasons and in line with the eligibility criteria defined in the protocol, but it introduces the possibility that relevant studies published in other languages were excluded. Given that digital health research is increasingly active in non-English-speaking regions, this restriction may have resulted in a somewhat skewed representation of the global evidence base. Similarly, the search did not systematically cover grey literature, conference proceedings, or preprint repositories, which means that recent or unpublished work that could have informed the synthesis was not captured. The final search was conducted on August 5, 2025, so studies published after that date are not reflected in these findings.

Although the study selection process was conducted by two independent reviewers with discrepancies resolved through discussion, any screening process of this kind involves a degree of interpretive judgment, particularly when determining whether a study's engagement with empathy was substantive enough to meet the eligibility criteria. Studies that addressed empathy only tangentially in their discussion sections were excluded, but the boundary between marginal and substantive engagement with the concept is not always sharp. It is plausible that a small number of relevant studies were excluded at the full-text screening stage for reasons that a different pair of reviewers might have decided differently.

5.2 Limitations of the Included Evidence Base

Perhaps the most consequential limitation of this review does not lie in its own methodology but in the nature of the studies it was able to include. The majority of the 40 studies in this corpus are short-term in design, with follow-up periods typically ranging from four to eight weeks. For a review focused on chronic disease management, where the relevant clinical question is whether chatbot support can be effective and acceptable over months and years rather than weeks, this is a significant gap. The four-month longitudinal study by ter Stal et al. (2021) is one of the few exceptions, and it was precisely this longer observation window that allowed the authors to detect the progressive decline in perceived empathy that shorter studies would have missed entirely. Studies conducted over shorter periods may systematically overestimate the relational quality and clinical promise of chatbot interventions by capturing initial novelty effects before they attenuate.

Related to this, the corpus shows a consistent and troubling pattern in terms of who participates in chatbot research. Most included studies recruited samples that were predominantly female, white, higher-educated, and based in high-income countries, with over 80 percent of the corpus originating from Europe, North America, and Australia. Wu et al. (2024) found that older adults and individuals with limited prior digital health experience exhibited substantially higher levels of AI hesitancy despite being among the most clinically relevant populations for chatbot-delivered chronic disease support. Espinoza et al. (2025) highlighted that digital literacy and infrastructure access shaped how users in Peru engaged with and benefited from their chatbot intervention in ways that would be invisible in studies conducted exclusively in high-resource settings. The populations most likely to benefit from accessible, low-cost digital health support are, in other words, the populations least represented in the research that has evaluated it.

Sample sizes across the corpus also varied considerably, with a number of feasibility and pilot studies recruiting fewer than 50 participants. While such studies serve a legitimate purpose in early-stage development, they are underpowered to detect anything but large effects and their findings cannot be generalized with confidence. Mitchell et al. (2021), for instance, noted that their comparison of chatbot and human coaching involved a Wizard-of-Oz prototype rather than a fully automated system, which limits how far their findings about relational differences can be applied to deployed chatbots. Several other studies involved prototypes tested under controlled conditions that may not reflect how users would engage with the same systems in real-world care environments.

Dropout and engagement attrition represent a further limitation of the included evidence. He et al. (2025) observed a progressive decline in engagement with the Roby smoking cessation chatbot across sessions that was not fully explained by clinical outcomes alone. Retention figures across the corpus range from approximately 70 to 88 percent, but attrition is rarely analyzed in relation to empathy perception or relational quality specifically. It is quite possible that users who disengaged from chatbot interventions did so precisely because they found the interactions insufficiently empathic or personally relevant, in which case the reported outcomes reflect the experiences of a self-selected group of more satisfied users and overstate the benefits for the broader population. This is a form of survivorship bias that the available data do not allow the review to address.

5.3 Conceptual and Synthetic Limitations

This review was designed as a scoping review rather than a systematic review with quantitative synthesis, which means it cannot establish effect sizes, does not pool data across studies, and cannot support causal claims about the relationship between specific empathic design features and particular outcomes. The findings are descriptive and interpretive rather than confirmatory. Where patterns are identified across the corpus, they reflect the qualitative judgment of the reviewers synthesizing a heterogeneous literature, and reasonable scholars might weight the same evidence differently and reach somewhat different conclusions.

A related limitation concerns the heterogeneity of outcome measures across the included studies. Because studies used such varied instruments to assess empathy, user experience, and clinical outcomes, direct comparison is rarely possible. The synthesis has had to work across these differences, grouping findings thematically rather than analytically, and this inevitably introduces some imprecision into the conclusions drawn. Where specific numbers are cited, they come from individual studies and should be understood in the context of that study's particular design and sample rather than as general estimates applicable across the field. In short, the review provides a map of the territory rather than a precise survey of it, and the distinctions that matter most for clinical practice and future research design will require the kind of targeted, well-powered, and methodologically consistent studies that the field has not yet consistently produced.

Chapter 6

Conclusion

The constraints identified in the preceding section do not invalidate the findings of this review, but they do ask for some caution in how those findings are read. Most of the studies in this corpus were short, recruited narrowly, and measured empathy in ways that cannot easily be compared to one another. These are genuine problems. What this review can still claim, despite them, is that it has produced the most systematic mapping of the available evidence on perceived empathy in chatbot interactions for chronic disease management attempted to date, and that the picture which emerges from that mapping is coherent enough to be useful.

The first research question asked what design strategies have been used to convey empathy in this context, and the answer is that they vary enormously, from isolated warm phrasing to fully theory-grounded relational architectures built around motivational interviewing or working alliance principles. That range is itself informative. It tells us that empathic design in this field has not yet standardized, that developers are making very different choices, and that the choices underpinned by clinical theory tend to produce more consistent results than those that rely on intuition alone. Sou et al. (2024) demonstrated that specific verbal cues can be experimentally tested and verified, which is an important step toward the kind of evidence base the field needs. But it remains the exception rather than the rule.

On how perceived empathy shapes outcomes, the evidence is encouraging but not straightforward. Several studies suggest that users can form something resembling a working alliance with a chatbot, report feeling supported and heard, and in some controlled designs show improvements in depression, anxiety, and treatment adherence alongside these relational gains. That is a meaningful pattern, though not yet a settled one. For people managing chronic conditions with limited access to regular clinical contact, a system that can provide consistent, non-judgmental, and personally responsive support fills a gap that would otherwise go unfilled. The question is not whether this is valuable but whether it can be sustained, and the longitudinal evidence reviewed here suggests it often cannot without a level of contextual adaptability that most current systems lack.

The third question, about moderating factors, produced perhaps the most practically important finding of the review. Perceived empathy is not a fixed property of a chatbot. It shifts depending on how long someone has been using the system, how well the system can track their changing situation, and increasingly, what other AI tools they have encountered before. The last of these is a relatively new problem, and it is worth stating plainly: chatbots that were considered adequately empathic before the widespread availability of large language models are now being rated as less empathic by users who have experienced those models firsthand. The standard is moving. Systems designed against the benchmarks of 2020 may already feel insufficient to users in 2025.

What this review cannot offer is a definitive account of how chatbots should be designed and deployed. The evidence is not yet strong enough for that. What it does offer is a reasonably clear map of where the field currently stands, what has been tried, what appears to work under what conditions, and where the most important gaps remain. The populations most likely to benefit from empathic chatbot support in chronic disease management are still the ones least represented in research. Measurement of empathy remains too fragmented to support cumulative knowledge-building. And the relational ceiling of current systems, the point beyond which automated interaction cannot provide what seriously ill patients actually need, has been identified but not yet adequately characterized.

These are solvable problems, or at least tractable ones. This corpus shows that empathy in chatbot interactions is not beyond the reach of design and evaluation, even if the tools for measuring it consistently are still underdeveloped, and that gives the field a starting point rather than a finished method. What it requires now is more rigorous methodology, longer studies, and a willingness to prioritize the patients who are hardest to reach over those who are easiest to recruit. That shift, more than any single technological advance, is probably what would make the most difference.

References

Anmella, G., Sanabra, M., Primé-Tous, M., Segú, X., Cavero, M., Morilla, I., Grande, I., Ruiz, V., Mas, A., Martín-Villalba, I., Caballo, A., Esteva, J. P., Rodríguez-Rey, A., Piazza, F., Valdesoiro, F. J., Rodriguez-Torrella, C., Espinosa, M., Virgili, G., Sorroche, C., ... Hidalgo-Mazzei, D. (2023). Vickybot, a chatbot for anxiety-depressive symptoms and work-related burnout in primary care and health care professionals: Development, feasibility, and potential effectiveness studies. *Journal of Medical Internet Research*, *25*, e43293. https://doi.org/10.2196/43293

Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology*, *8*(1), 19–32. https://doi.org/10.1080/1364557032000119616

Au, J., Falloon, C., Ravi, A., Ha, P., & Le, S. (2023). A beta-prototype chatbot for increasing health literacy of patients with decompensated cirrhosis: Usability study. *JMIR Human Factors*, *10*, e42506. https://doi.org/10.2196/42506

Bai, X., Wang, S., Zhao, Y., Feng, M., Ma, W., & Liu, X. (2025). Application of AI chatbot in responding to asynchronous text-based messages from patients with cancer: Comparative study. *Journal of Medical Internet Research*, *27*, e67462. https://doi.org/10.2196/67462

Baptista, S., Wadley, G., Bird, D., Oldenburg, B., Speight, J., & Group, T. M. D. C. R. (2020). Acceptability of an embodied conversational agent for type 2 diabetes self-management education and support via a smartphone app: Mixed methods study. *JMIR Mhealth Uhealth*, *8*(7), e17038. https://doi.org/10.2196/17038

Boustani, M., Lunn, S., Visser, U., & Lisetti, C. (2021). Development, feasibility, acceptability, and utility of an expressive speech-enabled digital health agent to deliver online, brief motivational interviewing for alcohol misuse: Descriptive study. *Journal of Medical Internet Research*, *23*(9), e25837. https://doi.org/10.2196/25837

Bruijnes, M., Kesteloo, M., Brinkman, & W.-P. (2023). Reducing social diabetes distress with a conversational agent support system: A three-week technology feasibility evaluation. *Frontiers in Digital Health*, *5*, 1149374. https://doi.org/10.3389/fdgth.2023.1149374

Coleman, S., Lynch, C., Worlikar, H., Kelly, E., Loveys, K., Simpkin, A. J., Walsh, J. C., Broadbent, E., Finucane, F. M., & O' Keeffe, D. (2025). "digital clinicians" performing obesity medication self-injection education: Feasibility randomized controlled trial. *JMIR Diabetes*, *10*, e63503. https://doi.org/10.2196/63503

Cook, D., Peters, D., Moradbakhti, L., Su, T., Da Re, M., Schuller, B. W., Quint, J., Wong, E., & Calvo, R. A. (2024). A text-based conversational agent for asthma support: Mixed-methods feasibility study. *Digital Health*, *10*, 1–18. https://doi.org/10.1177/20552076241258276

Dworkin, M. S., Lee, S., Chakraborty, A., Monahan, C., Hightow-Weidman, L., Garofalo, R., & Jimenez, A. (2019). Acceptability, feasibility, and preliminary efficacy of a theory-based relational embodied conversational agent mobile phone intervention to promote HIV medication adherence in young HIV-positive african american MSM. *AIDS Education and Prevention*, *31*(1), 17–37. https://doi.org/10.1521/aeap.2019.31.1.17

Easton, K., Potter, S., Bec, R., Bennion, M., Christensen, H., Grindell, C., Mirheidari, B., Weich, S., de Witte, L., Wolstenholme, D., & Hawley, M. S. (2019). A virtual agent to support individuals living with physical and mental comorbidities: Co-design and acceptability testing. *Journal of Medical Internet Research*, *21*(5), e12996. https://doi.org/10.2196/12996

Espinoza, F., Cook, D., Da Re, M., Cuba Fuentes, M. S., Butler, C. R., & Calvo, R. A. (2025). Designing AI-powered chatbots for dementia care in peru: Stakeholder engagement and field observations. *Interacting with Computers*, *2025*, 1–14. https://doi.org/10.1093/iwc/iwae057

Gomaa, S., Posey, J., Bashir, B., Basu Mallick, A., Vanderklok, E., Max Schnoll, M., Zhan, T., & Wen, K. Y. (2023). Feasibility of a text messaging–integrated and chatbot-interfaced self-management program for symptom control in patients with gastrointestinal cancer undergoing chemotherapy: Pilot mixed methods study. *JMIR Formative Research*, *7*, e46128. https://doi.org/10.2196/46128

Han, H. J., Mendu, S., Jaworski, B. K., Owen, J. E., & Abdullah, S. (2024). Assessing acceptance and feasibility of a conversational agent to support individuals living with post-traumatic stress disorder. *DIGITAL HEALTH*, *10*, 1–17. https://doi.org/10.1177/20552076241286133

Hasei, J., Hanzawa, M., Nagano, A., Maeda, N., Yoshida, S., Endo, M., Yokoyama, N., Ochi, M., Ishida, H., Katayama, H., Fujiwara, T., Nakata, E., Nakahara, R., Kunisada, T., Tsukahara, H., & Ozaki, T. (2025). Empowering pediatric, adolescent, and young adult patients with cancer utilizing generative AI chatbots to reduce psychological burden and enhance treatment engagement: A pilot study. *Frontiers in Digital Health*, *7*, 1543543. https://doi.org/10.3389/fdgth.2025.1543543

Hauser-Ulrich, S., Künzli, H., Meier-Peterhans, D., & Kowatsch, T. (2020). A smartphone-based health care chatbot to promote self-management of chronic pain (SELMA): Pilot randomized controlled trial. *JMIR Mhealth Uhealth*, *8*(4), e15806. https://doi.org/10.2196/15806

He, L., Basar, E., Wiers, R. W., Antheunis, M. L., & Krahmer, E. (2025). Chatting your way to quitting: A longitudinal exploration of smokers' interaction with a cessation chatbot. *Internet Interventions*, *39*, 100806. https://doi.org/10.1016/j.invent.2025.100806

Inkster, B., Kadaba, M., & Subramanian, V. (2023). Understanding the impact of an AI-enabled conversational agent mobile app on users' mental health and wellbeing with a self-reported maternal event: A mixed method real-world data mhealth study. *Frontiers in Global Women's Health*, *4*, 1084302. https://doi.org/10.3389/fgwh.2023.1084302

Inkster, B., Sarda, S., & Subramanian, V. (2018). An empathy-driven, conversational artificial intelligence agent (wysa) for digital mental well-being: Real-world data evaluation mixed-methods study. *JMIR Mhealth Uhealth*, *6*(11), e12106. https://doi.org/10.2196/12106

Lee, H. S., Wright, C., Ferranto, J., Buttimer, J., Palmer, C. E., Welchman, A., Mazor, K. M., Fisher, K. A., Smelson, D., O'Connor, L., Fahey, N., & Soni, A. (2025). Artificial intelligence conversational agents in mental health: Patients see potential, but prefer humans in the loop. *Frontiers in Psychiatry*, *15*, 1505024. https://doi.org/10.3389/fpsyt.2024.1505024

Leung, Y. W., So, J., Sidhu, A., Asokan, V., Gancarz, M., Gajjar, V. B., Patel, A., Li, J. M., Kwok, D., Nadler, M. B., et al. (2025). The extent to which artificial intelligence can help fulfill metastatic breast cancer patient healthcare needs: A mixed-methods study. *Current Oncology*, *32*(3), 145. https://doi.org/10.3390/curroncol32030145

Ludwig da Costa, F., Matzenbacher, L. S., Maia, I. S., Gheno, V., Brum, M. A. B., Barros, L. G. B., Blank, L. M., & Telo, G. H. (2025). "she has become my best friend": A qualitative study on the perspective of elderly with type 2 diabetes regarding the use of an interactive virtual assistant device for diabetes care and mental health promotion. *Acta Diabetologica*, *62*, 523–529. https://doi.org/10.1007/s00592-024-02377-z

Ma, Y., Achiche, S., Tu, G., Lessard, D., Engler, K., Lemire, B., Laymouna, M., de Pokomandy, A., Cox, J., & Lebouché, B. (2025). The first AI-based chatbot to promote HIV self-management: A mixed methods usability study. *HIV Medicine*, *26*(2), 184–206. https://doi.org/10.1111/hiv.13720

MacNeill, A. L., Doucet, S., & Luke, A. (2024). Effectiveness of a mental health chatbot for people with chronic diseases: Randomized controlled trial. *JMIR Formative Research*, *8*, e50025. https://doi.org/10.2196/50025

Martinez-Miranda, J., Martinez, A., Ramos, R., Aguilar, H., Jimenez, L., Arias, H., et al. (2019). Assessment of users' acceptability of a mobile-based embodied conversational agent for the prevention and detection of suicidal behaviour. *Journal of Medical Systems*, *43*, 246. https://doi.org/10.1007/s10916-019-1387-1

Mitchell, E. G., Maimone, R., Cassells, A., Tobin, J. N., Davidson, P., Smaldone, A. M., & Mamykina, L. (2021). Automated vs. human health coaching: Exploring participant and practitioner experiences. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1), 1–26. https://doi.org/10.1145/3449173

Peters, M. D. J., Godfrey, C., McInerney, P., Munn, Z., Tricco, A. C., & Khalil, H. (2020). Chapter 11: Scoping reviews. In E. Aromataris & Z. Munn (Eds.), *JBI manual for evidence synthesis*. Joanna Briggs Institute. https://doi.org/10.46658/JBIMES-20-12

Pienkowska, A., Ang, C.-S, Mammadova, M., Mahadzir, M. D. A., & Car, J. (2023). A diabetes education app for people living with type 2 diabetes: Co-design study. *JMIR Formative Research*, *7*, e45490. https://doi.org/10.2196/45490

Prochaska, J. J., Vogel, E. A., Chieng, A., Baiocchi, M., Maglalang, D. D., Pajarito, S., Weingardt, K. R., Darcy, A., & Robinson, A. (2021). A randomized controlled trial of a therapeutic relational agent for reducing substance misuse during the COVID-19 pandemic. *Drug and Alcohol Dependence*, *227*, 108986. https://doi.org/10.1016/j.drugalcdep.2021.108986

Prochaska, J. J., Vogel, E. A., Chieng, A., Kendra, M., Baiocchi, M., Pajarito, S., & Robinson, A. (2021). A therapeutic relational agent for reducing problematic substance use (woebot): Development and usability study. *Journal of Medical Internet Research*, *23*(3), e24850. https://doi.org/10.2196/24850

Rathnayaka, P., Mills, N., Burnett, D., De Silva, D., Alahakoon, D., & Gray, R. (2022). A mental health chatbot with cognitive skills for personalised behavioural activation and remote health monitoring. *Sensors*, *22*, 3653. https://doi.org/10.3390/s22103653

Sou, D., Kuhlmeier, F. O., & Kowatsch, T. (2024). Towards digital empathy in healthcare chatbots: A conceptual framework and empirical study. *Forty-Fifth International Conference on Information Systems (ICIS), Bangkok, Thailand*.

Tan, T. C., Roslan, N. E. B., Li, J. W., Zou, X., Chen, X., Ratnasari, & Santosa, A. (2023). Patient acceptability of symptom screening and patient education using a chatbot for autoimmune inflammatory diseases: Survey study. *JMIR Formative Research*, *7*, e49239. https://doi.org/10.2196/49239

ter Stal, S., Sloots, J., Ramlal, A., op den Akker, H., Lenferink, A., & Tabak, M. (2021). An embodied conversational agent in an ehealth self-management intervention for chronic obstructive pulmonary disease and chronic heart failure: Exploratory study in a real-life setting. *JMIR Human Factors*, *8*(4), e24110. https://doi.org/10.2196/24110

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., et al. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine*, *169*(7), 467–473. https://doi.org/10.7326/M18-0850

van Bussel, M., Oud, G., & Jansen, M. (2022). Patient acceptance of virtual assistants in radiation oncology: A mixed-methods study. *BMC Health Services Research*, *22*, 890. https://doi.org/10.1186/s12913-022-08189-7

Wu, P. F., Summers, C., Panesar, A., Kaura, A., & Zhang, L. (2024). AI hesitancy and acceptability—perceptions of AI chatbots for chronic health management and long COVID support: Survey study. *JMIR Human Factors*, *11*, e51086. https://doi.org/10.2196/51086

Yan, Z., Liu, J., Fan, Y., Lu, S., Xu, D., Yang, Y., Wang, H., Mao, J., Tseng, H.-C, Chang, T.-H, & Chen, Y. (2025). Ability of ChatGPT to replace doctors in patient education: Cross-sectional comparative analysis of inflammatory bowel disease. *Journal of Medical Internet Research*, *27*, e62857. https://doi.org/10.2196/62857

Zhu, J., Dong, A., Wang, C., Veldhuizen, S., Abdelwahab, M., Andrew Brown, A., Selby, P., & Rose, J. (2025). The impact of ChatGPT exposure on user interactions with a motivational interviewing chatbot: Quasi-experimental study. *JMIR Formative Research*, *9*, e56973. https://doi.org/10.2196/56973

Appendix A

Full Search Strategies Across All Databases

A.1 PubMed

```
((chatbot*[Title/Abstract] OR "conversational agent*"[Title/Abstract] OR "
    social bot*"[Title/Abstract] OR softbot*[Title/Abstract] OR "virtual
    agent*"[Title/Abstract] OR "automated agent*"[Title/Abstract] OR "
    automated bot*"[Title/Abstract] OR "virtual therap*"[Title/Abstract] OR
     "virtual assistant*"[Title/Abstract]) AND ("chronic disease*"[Title/
    Abstract] OR "chronic condition*"[Title/Abstract] OR "long-term
    condition*"[Title/Abstract] OR "disease management"[Title/Abstract] OR
    "healthcare"[Title/Abstract] OR "digital health"[Title/Abstract] OR "
    health management" / chronic patient*[Title/Abstract] OR chronic user*[
    Title/Abstract] OR "chronic condition*"[Title/Abstract] OR "disease
    management"[Title/Abstract] OR health*[Title/Abstract])) AND (empath*[
    Title/Abstract] OR "user experience*"[Title/Abstract] OR perception*[
    Title/Abstract] OR attitude*[Title/Abstract] OR "emotional intelligence
    "[Title/Abstract] OR "affective computing"[Title/Abstract] OR "
    relational agent*"[Title/Abstract] OR "emotional support"[Title/
    Abstract])
```

A.2 Scopus

```
Query = ( TITLE-ABS-KEY ( chatbot* OR "conversational agent*" OR "social
    bot*" OR softbot* OR "virtual agent*" OR "automated agent*" OR "
    automated bot*" OR "virtual therap*" OR "virtual assistant*" ) AND
    TITLE-ABS-KEY ( "chronic disease*" OR "chronic condition*" OR "long-
    term condition*" OR "disease management" OR "healthcare" OR "digital
    health" OR "health management" / AND chronic AND patient* OR chronic
    AND user* OR "chronic condition*" OR "disease management" OR health* )
    AND TITLE-ABS-KEY ( empath* OR "user experience*" OR perception* OR
    attitude* OR "emotional intelligence" OR "affective computing" OR "
    relational agent*" OR "emotional support" ) )
```

A.3 Web of Science

```
((AB=(chatbot* OR "conversational agent*" OR "social bot*" OR softbot* OR "
    virtual agent*" OR "automated agent*" OR "automated bot*" OR "virtual
    therap*" OR "virtual assistant*" )) AND AB=("chronic disease*" OR "
    chronic condition*" OR "long-term condition*" OR "disease management"
    OR "healthcare" OR "digital health" OR "health management" / chronic
    patient* OR chronic user* OR "chronic condition*" OR "disease
    management" OR health*)) AND AB=(empath* OR "user experience*" OR
    perception* OR attitude* OR "emotional intelligence" OR "affective
    computing" OR "relational agent*" OR "emotional support")Results = 611
```

A.4 PsycINFO

```
Query = AB (chatbot* OR "conversational agent*" OR "social bot*" OR softbot
    * OR "virtual agent*" OR "automated agent*" OR "automated bot*" OR "
    virtual therap*" OR "virtual assistant*") AND AB ("chronic disease*" OR
     "chronic condition*" OR "long-term condition*" OR "disease management"
     OR "healthcare" OR "digital health" OR "health management" / chronic
    patient* OR chronic user* OR "chronic condition*" OR "disease
    management" OR health*) AND AB (empath* OR "user experience*" OR
    perception* OR attitude* OR "emotional intelligence" OR "affective
    computing" OR "relational agent*" OR "emotional support")
```

A.5 CINAHL

```
Query = AB (chatbot* OR "conversational agent*" OR "social bot*" OR softbot
    * OR "virtual agent*" OR "automated agent*" OR "automated bot*" OR "
    virtual therap*" OR "virtual assistant*") AND AB ("chronic disease*" OR
     "chronic condition*" OR "long-term condition*" OR "disease management"
     OR "healthcare" OR "digital health" OR "health management" / chronic
    patient* OR chronic user* OR "chronic condition*" OR "disease
    management" OR health*) AND AB (empath* OR "user experience*" OR
    perception* OR attitude* OR "emotional intelligence" OR "affective
    computing" OR "relational agent*" OR "emotional support")
```