JMIR Biomedical Engineering

Engineering for health technologies, medical devices, and innovative medical treatments and procedures
Volume 11 (2026) ISSN 2561-3278 Editor in Chief: Javad Sarvestan, PhD

Contents

Original Papers

Conceptual Model for the Integration of Marketing Strategies and Biomedical Innovation in Patient-Centered Care: Mixed Methods Study

Atantra Das Gupta[1], BPharm, PGDBM, MS, MEng; Yashpal Yadav[2], BEng, MEng, PhD

[1]Department of Biomedical Engineering, Nims University Rajasthan, NH-11C Delhi-Jaipur Highway, Jaipur, India

[2]Department of Electrical, Electronics & Electric Vehicle Engineering, Nims University Rajasthan, Jaipur, India

Corresponding Author:
Atantra Das Gupta, BPharm, PGDBM, MS, MEng
Department of Biomedical Engineering, Nims University Rajasthan, NH-11C Delhi-Jaipur Highway, Jaipur, India

Abstract

Background: The increasing integration of biomedical technology and digital marketing is quickly transforming how patients engage with health care. The patient as an organization (PAO) model is explored in this study. The PAO model encourages patients to be active participants in health care decisions by leveraging wearables, mobile health (mHealth) apps, artificial intelligence (AI) platforms, and health care marketing strategies.

Objective: This study aims to examine how the PAO model is evolving in practice and gain insight into both the opportunities and challenges created by the intersection of biomedical innovation and marketing practices in patient care.

Methods: The scoping review was conducted across Scopus, Web of Science, PubMed, and Google Scholar. Selection criteria included articles published from 2014 to 2024. Studies were included if they examined connections among biomedical technologies, marketing strategies, and models of behavior and organizations. Studies lacking interdisciplinary focus or methodological rigor were excluded. Since this work was exploratory, it did not require a strict bias assessment. Additionally, findings derived from qualitative analysis of 18 semistructured interviews with patients, health care professionals, and technologists regarding their experiences with digital technologies and perceptions of trust, autonomy, and engagement were analyzed. Thematic analysis was applied to these interviews using open, axial, and selective coding.

Results: From an initial pool of 22,740 records, 45 studies met the inclusion criteria and were analyzed. The review revealed that the integration of AI-based personalization, biosensors, and remote monitoring with marketing strategies, such as segmentation, customer relationship management systems, and behavioral nudging, offers potential to enhance patient autonomy and engagement. However, most studies were descriptive or exploratory, with limited empirical evaluation, particularly regarding ethical risks and digital inequality. Qualitative findings further illustrated how patients are adopting organizational behaviors, such as self-monitoring, real-time decision-making, and strategic management of health data. The following 5 key themes emerged: (1) patients as autonomous digital actors, (2) digital health as a behavioral ecosystem, (3) inequities in digital empowerment, (4) negotiating trust and ethical transparency, and (5) blended care as the preferred future. Although many participants embraced digital tools, concerns about data transparency, algorithmic bias, and loss of human connection highlighted important barriers to equitable adoption.

Conclusions: The PAO model shows strong potential for personalizing care and engaging patients in health care. However, it is important to note that, so far, conceptual models have dominated the PAO literature, with little empirical evidence to support them. Therefore, as health care practices increasingly integrate digital technologies, it is crucial to develop appropriate safeguards for PAO models.

(JMIR Biomed Eng 2026;11:e77115) doi:10.2196/77115

KEYWORDS

biomedical technology; patient engagement; digital health; AI in health care; health marketing; wearables; CRM; customer relationship management

Introduction

The concept of the patient as an organization (PAO) marks a significant shift in digital health care, redefining the patient as an active participant, strategic decision-maker, and key stakeholder in their care. Instead of being passive recipients of treatment, patients are increasingly managing their health data, engaging with providers, and shaping the design of the health care system. Based on organizational theory and health care strategy, this model encourages patients to take on roles typically held by structured entities, emphasizing self-management, participation, and governance. The World Health Organization (WHO) has recognized this change, calling for greater patient involvement in the development and implementation of health care systems to promote more responsive and effective care [1].

Despite its increasing importance, the PAO model remains mainly theoretical. Although benefits such as better health outcomes, lower system costs, and increased patient satisfaction are often cited, most practical efforts focus on mobile apps and wearable devices. However, truly decentralized health care, another major shift, depends on broader technological integration. Future systems will need to incorporate artificial intelligence (AI)–powered diagnostics, big data analysis, home-connected medical devices, and generative AI platforms such as ChatGPT [2]. These tools are not just extras; they change how health care is delivered, expanded, and customized.

Currently, patient-centered care tends to focus heavily on disease management. However, health care should encompass more than that: It should include wellness promotion, behavioral support, and preventive care [3]. This broader approach reflects a shift from reactive, "disease-care" models toward proactive, wellness-oriented digital health systems. The expanding role of mobile and digital platforms in areas like fitness tracking, lifestyle coaching, and preventive screening reflects this growth, creating not only new models of care but also new opportunities for health business innovation [4].

Within the PAO framework, the patient becomes a digitally connected and ethically engaged actor, actively managing their health records, co-designing service delivery, and even contributing to policymaking and research. This reframing introduces new dimensions of trust, transparency, and autonomy. It also brings the patient experience closer to the structure of advocacy-driven nonprofit organizations that represent patient and caregiver interests. However, trust is not simply a desirable outcome; it is essential, and yet, the ethical dimensions of digital health marketing, including privacy, consent, and algorithmic bias, are often underexplored [5].

The adoption of technology in health care has grown rapidly. More than 50% of patients now use telemedicine, more than 90% of care providers utilize electronic health records, and digital platforms such as social media are commonly used for health communication. However, these advances often mask ongoing digital inequalities along lines of race, geography, income, and education [6]. To address this, health care organizations should learn from business, particularly in behavioral segmentation, predictive analytics, and customer relationship management (CRM). Strategic planning and customized communication are crucial for expanding access, increasing reach, and improving health outcomes.

At the heart of this transformation lies biomedical technology: a fusion of biology, engineering, and computing designed to enhance care across the continuum. AI-powered imaging tools, biosensors, implantable monitors, and smart prosthetics now enable real-time diagnostics, adaptive treatment, and precision health management [7]. Examples include robotic-assisted surgeries that reduce risk, insulin pumps that automatically respond to glucose fluctuations, and wearable devices that track behavior and symptoms in real time, bridging gaps in traditional health care.

These technologies demand parallel evolution in data governance, security, and ethics. Management leaders must ensure that AI-powered systems comply with privacy regulations, cybersecurity frameworks, and inclusive design principles. The WHO's *Global Strategy on Digital Health 2020‐2025* reinforces this direction, advocating for equity-focused digital transformation, especially in resource-constrained regions.

The PAO model is a convergence point, strategically integrating biomedical technology, behavioral science, and health care marketing. It supports a move from static, episodic treatment to dynamic, data-informed, and personalized health care management. Patients are no longer seen simply as users of care; they are empowered collaborators who coproduce health outcomes through technology-enabled engagement and decision-making [8].

Marketing frameworks provide a valuable lens for translating innovation into actionable steps. The traditional 4 Ps of product, price, place, and promotion take on renewed significance in the digital health era: [9]

- Product includes AI diagnostics, wearable biosensors, robotic interventions, and mobile point-of-care tools that support accuracy, personalization, and autonomy [10].
- Price is reflected in value-based care models like pay-for-performance, which reward quality and efficiency enabled through biomedical monitoring [11].
- Place reflects that care is no longer confined to clinical spaces. With portable, internet-connected tools, services can reach patients in their homes, remote regions, or emergency settings [6].
- Promotion through digital communication, including ethically designed AI messaging, social media campaigns, and CRM outreach, ensures that patients receive accurate, timely, and personalized health information [12].

The convergence of digital health technologies with health care delivery not only drives innovation but also supports global health equity. Scalable tools, like mobile diagnostics and cloud-based platforms, can extend care to underserved communities and help reduce disparities.

Although digital tools offer benefits such as improved access, personalized communication, and behavior change, deeper issues, like digital exclusion, ethical concerns, and systemic barriers, remain underexplored. Challenges such as unequal access, low digital literacy, and lack of trust persist, particularly in marginalized populations.

To address these gaps, this study combined a scoping literature review with qualitative research to examine the evolving concept of the PAO. We explored how patients increasingly engage in organizational-like behaviors, such as self-tracking, strategic participation, and co-creating care, while facing barriers related to equity, ethics, and infrastructure.

By grounding the PAO model in interdisciplinary and empirical research, we moved beyond theory to examine how emerging technologies, like AI tools, wearables, and mobile health (mHealth) apps, are reshaping patient roles. Our goal was to understand how these tools influence patient engagement, decision-making, and autonomy within connected, data-driven health care systems.

Methods

Mixed Methods Design

This study followed a mixed methods design consisting of 2 stages: Stage 1 involved a scoping review of the literature, and stage 2 included a qualitative study using semistructured interviews.

The study design was guided by the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) framework and informed by constructivist-grounded theory principles to allow for the emergence of themes grounded in real-world patient experience.

Stage 1: Scoping Review Method

Design and Framework

This study used a scoping review methodology guided by the PRISMA-ScR framework to ensure clarity, transparency, and academic rigor [13]. This review examined how biomedical technologies and marketing theory intersect within the emerging PAO model. Using a scoping review approach, we mapped contributions across health care, behavioral science, and management to highlight key insights and gaps. This helped build a clearer picture of how digital tools and health marketing can together support more responsive, data-driven, and patient-centered care.

Search Strategy and Data Sources

To achieve disciplinary depth and interdisciplinary breadth, an exhaustive academic search was conducted across 4 prominent databases: Scopus, Web of Science, PubMed, and Google Scholar. Boolean operators and carefully selected keyword combinations were used to explore the intersections between digital health, biomedical innovation, and health care marketing.

Key terms included *healthcare*, *CRM (Customer Relationship Management)*, *biomedical innovation*, *AI personalization in healthcare*, *digital nudging*, *behavioral economics in healthcare*, *health belief model*, *segmentation*, and *patient targeting*.

Research Questions

Research question (RQ) 1 was "In what ways do patients use biomedical technologies—such as wearables, mHealth apps, and AI-powered tools—to take control of their health and engage in self-management similar to organizational behavior, and how do ethical safeguards and equitable access shape these practices across diverse populations?"

RQ2 was "How are marketing strategies such as personalization, segmentation, and CRM integrated into biomedical technologies to enhance patient engagement, treatment adherence, and behavior change, and what equity and ethical challenges arise in this integration?"

RQ3 was "What social, structural, and contextual factors, such as digital literacy, access to infrastructure, and socioeconomic status, affect patients' ability to use biomedical technologies effectively and equitably within the PAO framework?"

RQ4 was "How do patients understand and respond to ethical concerns associated with biomedical technologies, including data-driven personalization, algorithmic decision-making, and digital nudging, and how do these perceptions influence trust, autonomy, and willingness to adopt such tools?"

RQ5 was "How is the model of the PAO delivered by way of biomedical technologies, and how can it be conceived and regulated to give priority to equity, inclusiveness, and ethical integrity as core conditions for success?"

Together, these findings indicated that, although the literature offers rich descriptive themes, it often lacks rigorous appraisal of effectiveness and equity, which weakens the PAO model's empirical foundation. This limitation informed the qualitative phase of this study, designed to provide deeper, evidence-based insights.

Describing How the Scoping Review Informed the Qualitative Phase (From Stage 1 to Stage 2)

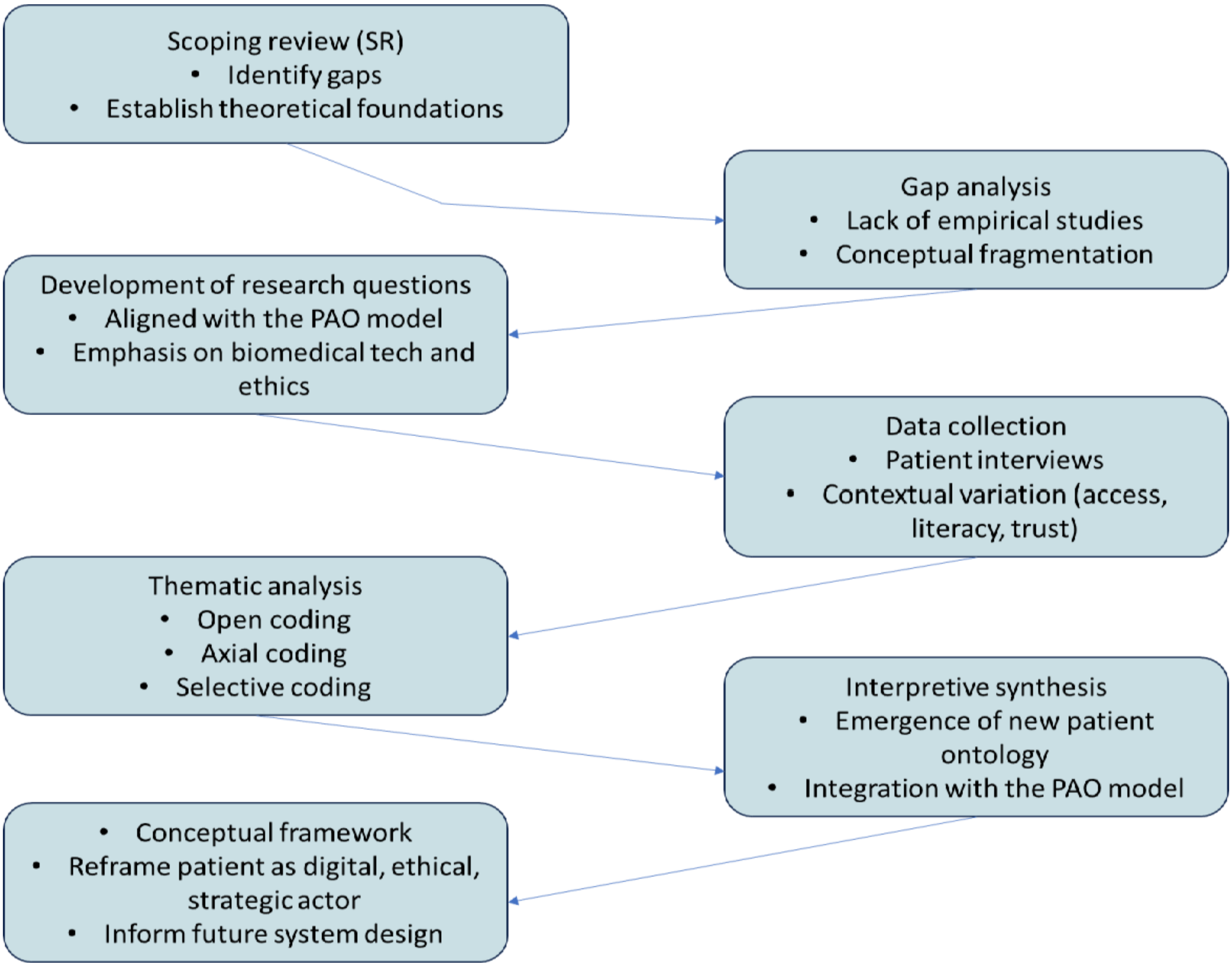
Overview

Although the scoping review offered a strong conceptual foundation, highlighting how digital tools and marketing strategies are shaping a new model of patient engagement, it also revealed important gaps. Many studies were exploratory and lacked insight into how patients in real life experience these innovations. To address this, we turned to the second phase of the study: qualitative interviews. This phase aimed to ground the theoretical potential of the PAO model in the voices and lived experiences of patients, clinicians, and digital health developers. By listening closely to how people interact with technologies in their everyday health routines, we were able to explore how the PAO model is beginning to take shape beyond theory and where its promises meet real-world complexity.

Research Design

Figure 1 illustrates a visual roadmap linking the scoping review to the qualitative phase of this study. The scoping review provided a foundation by highlighting key conceptual gaps, such as the lack of empirical evidence and fragmented definitions of the PAO. These insights directly guided the development of RQs centered on biomedical technology, ethics, and patient empowerment. Based on these questions, qualitative data were collected through in-depth patient interviews, focusing on real-world factors like digital access, literacy, and trust. Using a structured thematic analysis process, including open, axial, and selective coding, patterns that emerged were combined into a coherent framework. This integration of findings not only refined the PAO model but also helped develop a new patient ontology that portrays individuals as active, ethical, and strategic contributors to the future design of health systems.

Figure 1. Methodological flowchart for developing and synthesizing the patient as an organization (PAO) conceptual framework based on the authors' scoping synthesis and principles of qualitative research design.



Stage 2: Qualitative Study Method

Overview

The qualitative study aimed to explore how patients, health care professionals, and biomedical technologists experience the integration of digital tools and marketing strategies in health care. Although the sample size (n=18) may seem small, it was intentionally chosen to align with the study's exploratory goals and to prioritize conceptual depth and theoretical insight. Participants were purposively selected from 3 stakeholder groups to ensure diversity in experience, role, and digital exposure. Saturation was reached after 12 interviews, with additional participants confirming the existing themes rather than adding new ones. Beyond saturation, the adequacy of the sample is supported by the richness and variation in responses, which provided enough depth to identify recurring themes related to autonomy, trust, digital exclusion, and the changing role of the patient. The study prioritized analytical depth over statistical breadth, aligning with qualitative methods that focus on developing conceptual frameworks rather than producing generalizable results. Furthermore, because the study's goal was to refine the emerging PAO model and examine its real-world applications, this approach enabled a layered, interpretive analysis of how digital tools influence patient behavior and engagement. Nonetheless, broader claims about the model's applicability will require further empirical testing across larger, more diverse populations.

Qualitative Research Design

In today's evolving health care ecosystem, the PAO concept lies at the center of transformation, shaped by advancements in biomedical technologies and strategic health care marketing approaches [14]. Marketing plays a crucial role in influencing how patients perceive, adopt, and integrate such technologies into their daily self-care routines [15]. However, adoption remains a nuanced process, shaped by factors such as technological trust, usability, privacy concerns, digital literacy, and the demand for accessible and personalized solutions.

This qualitative study was designed to explore the lived experiences of patients, health care providers, and biomedical technologists to understand the interplay between patient empowerment, biomedical innovation, and health care marketing strategies. Specifically, it sought to examine the facilitators and barriers to adoption of digital health technologies and identify how marketing can be aligned with the values and expectations of technologically enabled, self-managing patient communities.

Participants and Sampling

Participants were recruited using purposive sampling to ensure that they were relevant to the research focus. The inclusion criteria required participants to have a minimum of 1 year of experience using or developing digital health technologies, such as wearable biosensors, mHealth apps, or AI-powered tools. Participants represented 3 stakeholder groups: patients actively using digital tools, health care professionals implementing these technologies, and biomedical technologists involved in the design and deployment of these technologies. Individuals with no relevant experience or those unable to provide informed consent were excluded from the study.

Data Collection

We conducted 18 semistructured interviews using secure online video conferencing platforms to facilitate accessibility and geographic diversity. Each interview lasted between 45 minutes and 60 minutes and was audio-recorded with participant consent. The interview protocol included open-ended questions designed to elicit insights into technology adoption behaviors, trust dynamics, usability experiences, marketing communication preferences, and ethical concerns.

Data Analysis

The study used a grounded theory–inspired coding process, comprising open, axial, and selective coding, to identify patterns and themes within the data [16]. Interviews were transcribed verbatim and analyzed iteratively using qualitative analysis software. Thematic saturation was reached after 12 interviews, at which point adding new data no longer yielded novel insights. Consistent with the *Law of Diminishing Returns in Qualitative Research* [17], data collection ceased at this point to maintain methodological efficiency and thematic clarity.

This approach ensured that each interview contributed meaningfully to understanding how patient organizations engage with digital tools, what shapes their decision-making, and how biomedical solutions can be better aligned with trust, values, and behavioral drivers.

Ethical Considerations

The study was conducted in accordance with ethical guidelines for qualitative research. All participants provided informed consent. Data were anonymized, securely stored, and used solely for research purposes. Because the interviews were conducted for academic purposes only and will not be used for commercial or promotional purposes, review by an institutional review board was not required [18-20].

Results

Stage 1: Scoping Review Results

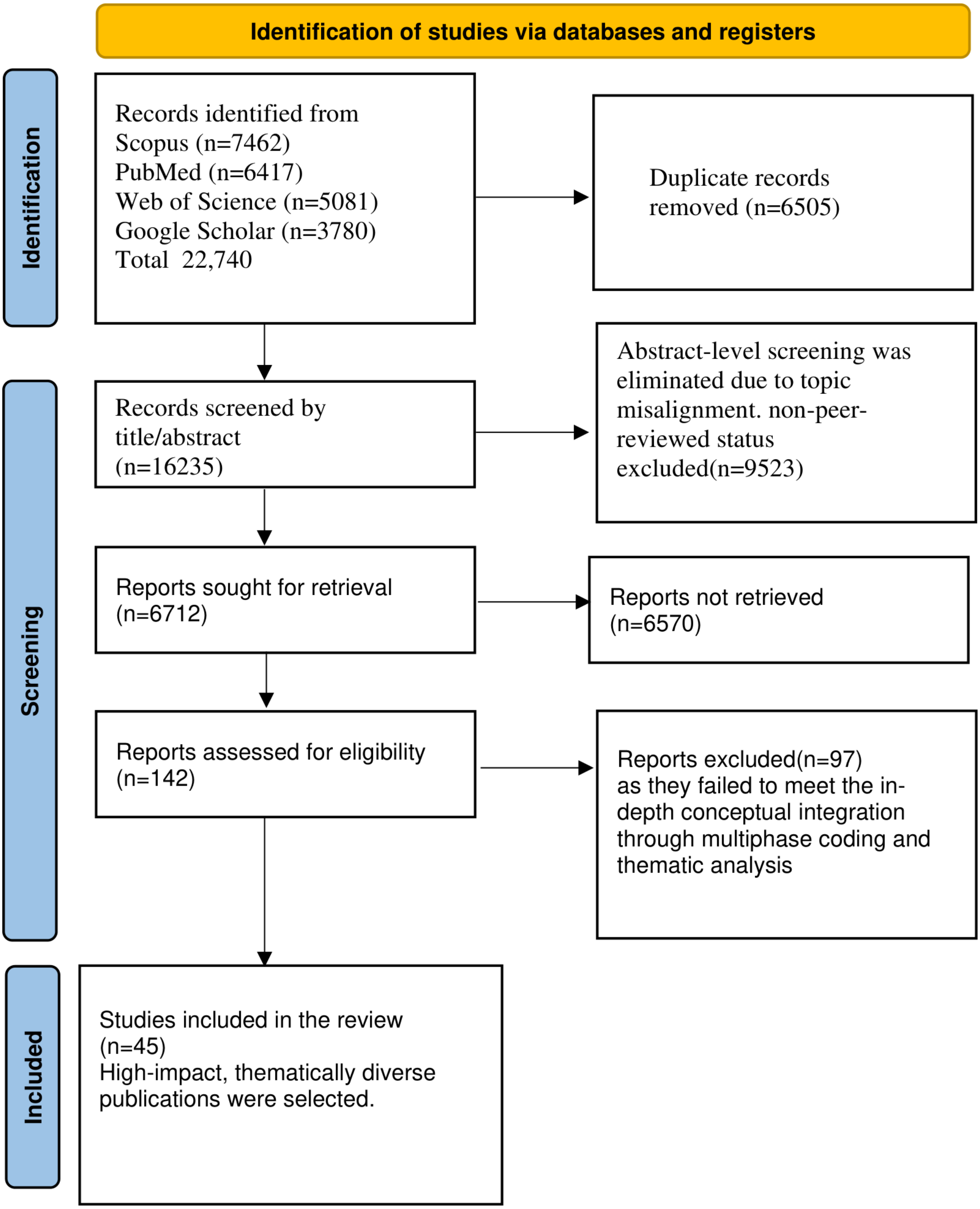
Mapping the PAO Landscape

The search process yielded a total of 22,740 records (Scopus: n=7462; Web of Science: n=5081; PubMed: n=6417; and Google Scholar: n=3780). Following the removal of 6505 duplicate entries, 16,235 unique records remained for initial screening. Abstract-level review excluded 9523 studies due to topic misalignment, lack of peer review, or language barriers. The remaining 6712 full-text articles were assessed for methodological quality and thematic alignment.

Of these, 142 studies met the inclusion criteria for the qualitative synthesis. However, 97 were excluded during deeper analysis because they lacked the conceptual integration required by the multiphase coding framework. Ultimately, 45 high-impact and thematically diverse publications were selected for inclusion.

This entire process adhered to the PRISMA-ScR guidelines and is visually summarized in Figure 2, which outlines the identification, screening, eligibility, and inclusion phases of the review.

Figure 2. PRISMA-ScR (Preferred Reporting Items for Scoping Reviews and Meta-Analyses Extension for Scoping Reviews) flow diagram [21].



Scoping Review Findings: Mapping Biomedical and Marketing Integration Within the PAO Model

From over 22,000 records, the scoping review identified 45 key studies exploring how biomedical innovation intersects with health care marketing under the PAO model. Common themes included AI-driven segmentation, behavioral economics, CRM, and digital nudging, highlighting new ways to personalize care and boost patient engagement. Many studies also pointed to the growing use of wearables, biosensors, and mHealth platforms to support real-time feedback and predictive analytics.

However, the evidence was uneven. Most research was exploratory, based on small pilot studies or case studies, with few rigorous evaluations or long-term outcomes. Promising tools such as AI segmentation and CRM often failed to account for cultural, ethical, or trust-related factors. Digital nudging raised concerns around autonomy, but few studies tested its real-world impact.

In short, although the PAO model shows strong potential, its practical value is still emerging. Clearer evidence and stronger ethical frameworks are needed to turn these innovations into scalable, patient-centered solutions.

Three Core Themes Emerged Across the Literature

AI-Driven Personalization and Segmentation

Many studies examined how AI and predictive analytics are used to tailor interventions to individual health profiles. These approaches mirror commercial segmentation strategies, enabling more responsive, targeted care.

CRM and Engagement

The adaptation of CRM systems in health care is enabling more continuous and personalized communication between patients and providers. This has shown promise in improving satisfaction, adherence, and long-term engagement.

Digital Nudging and Behavioral Influence

Several articles discussed the use of behavioral nudges embedded in apps and digital platforms to encourage healthier choices. However, concerns around autonomy and the ethical boundaries of persuasive design were rarely explored in depth.

Gaps in the Evidence Base

In addition to these thematic strengths, the review also highlighted important gaps in the current evidence base. Most studies were exploratory, drawing on pilot projects, descriptive analyses, or commercial analogies rather than longitudinal evaluations or controlled trials. Empirical testing of effectiveness, ethical risks, and digital equity was limited, particularly regarding diverse patient representation and co-designed solutions. Issues like data transparency, algorithmic bias, and unequal access to technology received minimal attention.

Despite these limitations, the review provided a rich conceptual foundation for understanding how the PAO model is developing. It also helped guide the qualitative phase of the study by identifying key mechanisms such as personalization, behavioral design, and engagement technologies through which patients are increasingly acting as strategic, data-informed participants in their care.

Stage 2: Qualitative Findings—Patients as Self-Organizing Actors

Participant Overview

Interviews took place with 18 participants who had direct and indirect experience with patients and represented clinicians and digital health technology experts. The participants had various experiences with technologies such as AI-based health applications and patient portals.

Exploring Patient Experiences With Biomedical Technologies Through the PAO Lens

Table 1 shows how patient experiences and stakeholder insights informed the conceptual development of the PAO model across behavioral, ethical, structural, and systemic dimensions by mapping RQ1-RQ5 to the thematic categories that emerged during the qualitative analysis.

Table . Alignment of interview themes with corresponding research questions (RQs) in the patient as an organization (PAO) framework, developed by the authors based on thematic analysis of qualitative interview data.

Interview question	Related theme
RQ1	Patients as autonomous digital actors
RQ2	Digital health as a behavioral ecosystem
RQ3	Inequities in digital empowerment
RQ4	Trust and ethical transparency
RQ5	Blended care and systems-level PAO framework

How Do Patients Describe Their Experiences With Biomedical Technologies (eg, Wearables, Health Apps, AI Tools) for Monitoring, Managing, and Making Decisions About Their Health?

This explored self-regulation, autonomy, and organizational behaviors from the patient's perspective. The interview focus was on "How do you use digital tools to track or manage your health? What role do these tools play in your decision-making?"

In What Ways Do Patients Perceive That Digital Health Platforms Apply Personalization, Nudging, or Targeted Communication Strategies to Influence Their Health Behaviors?

This investigated how marketing strategies are experienced through biomedical technology (eg, STP, CRM, and behavioral nudging). The interview focus was on "Do your apps or tools provide personalized suggestions or reminders? How do these affect your motivation or trust?"

What Challenges Do Patients Face With Accessing, Understanding, and Effectively Using Biomedical Technologies, Particularly Across Socioeconomic or Geographic Contexts?

This addressed the digital divide, equity, and structural limitations to PAO operationalization. The interview focus was on "What makes it easy or difficult for you to use digital health tools? How do factors like internet access or digital literacy impact your experience?"

How Do Patients View Consent, Data Privacy, and Transparency Issues When Interacting With AI-Powered or Data-Driven Biomedical Tools?

This explored ethical concerns tied to trust, data handling, and algorithmic personalization. The interview focus was on "How do your apps or devices use your health data? How does this affect your trust in the system?"

How Do Patients Envision the Ideal Balance Between Digital Technology and Human Interaction in Health Care, and What Features Do They Believe a Future Digital Health System Should Include?

This aimed to co-create or inform a future interdisciplinary PAO framework. The interview focus was on "How do you feel about relying on technology versus speaking with a health care provider? What would your ideal digital health system look like?"

Understanding the Operationalization of the PAO Model

Thematic Analysis

Thematic analysis of the qualitative interviews uncovered a set of interconnected themes that demonstrate how patients are increasingly adopting organizational-like roles within digital health ecosystems [22]. These themes corresponded to the core pillars of the PAO framework, specifically behavioral agency, digital engagement, structural equity, ethical trust, and systems-level integration.

As shown in Table 2, each theme was closely aligned with one of the core RQs (RQ1–RQ5), demonstrating how participants' lived experiences reflect the evolving roles of patients as autonomous decision-makers, digital collaborators, and ethical stakeholders. From personalized technology use and behavioral self-regulation to trust in AI-driven tools and systemic challenges in access, the findings provide a nuanced understanding of how the PAO model is being enacted in real-world contexts.

Table . Thematic evaluation of biomedical technology within the patient as an organization (PAO) framework based on a thematic synthesis of selected literature in the scoping literature review by the authors.

Theme	Key studies (authors, year)	Role of biomedical technology	Critical evaluation	Thematic transition
Segmentation, targeting, and personalization (STP)	Brommels, 2020 [23]	AI[a]-driven segmentation tools and digital health communication platforms support tailored interventions.	Although conceptually strong, most evidence comes from small-scale or single-site pilots. Limited comparative testing undermines PAO's claim to broad applicability and limits scalability.	It provides the foundation for targeted interventions but requires stronger, cross-system validation to serve as a reliable PAO mechanism.
Health belief model (HBM)	Ahadzadeh et al, 2015 [24]	AI-powered symptom checkers and telemedicine platforms tailor communication to patient risk perceptions.	Evidence is effective for literate and digitally fluent users but neglects populations with low health literacy or limited digital access. This exclusion undermines PAO's inclusivity.	It establishes a psychological basis for personalization, but without testing in vulnerable groups, it risks reinforcing inequities within PAO.
Behavioral influence and social marketing	Evans, 2006 [25]	Behavioral analytics and health apps are designed for broad population-level engagement.	This demonstrates strong public health influence, yet applications in chronic care and low-resource contexts remain underexplored. The lack of contextual adaptation limits PAO's scalability.	It informs engagement strategies but remains descriptive; long-term effectiveness must be empirically tested for PAO adoption.
Customer relationship management (CRM)	Mohiuddin, 2019 [26]	Predictive communication tools, EHR[b]-linked messaging, and reminder systems sustain engagement.	Although these are effective for engagement, the evidence draws heavily on commercial analogies. Few longitudinal health care studies exist, leaving CRM's role in PAO sustainability unproven.	It suggests potential for trust-building and continuity but risks oversimplifying health care relationships unless tested in diverse care environments.
Branding and trust-building	Mohamed, 2022 [27]	Transparent design interfaces, ethical AI systems, and privacy protocols support trust.	Although this is conceptually robust, most studies are cross-sectional, offering little insight into how trust evolves. This gap weakens PAO's ethical foundation.	It establishes an ethical entry point for PAO adoption, but without longitudinal studies, it remains more aspirational than practical.
Innovation adoption (diffusion theory)	Dearing and Cox, 2018 [28]	Wearables, telehealth tools, and peer-based adoption stories encourage uptake.	It explains early adoption effectively but overlooks structural barriers for marginalized groups. Evidence is skewed toward digitally privileged populations.	This drives momentum for mainstreaming PAO but requires inclusive adoption models to ensure equity.
Behavioral nudging and economics	Auf et al, 2021 [29]	Gamification, default settings, and subtle interface nudges are embedded in health apps.	These encourage short-term behavior change, yet few real-world studies examine ethical limits in high-stakes care. Weak empirical grounding risks compromising autonomy in PAO.	This supports digital habit formation but needs stronger ethical evaluation to avoid coercion in PAO practices.
Information-motivation-behavioral skills (IMB) model	Rongkavilit et al, 2010 [30]	Decision aids, chatbots, and adaptive mobile learning systems build skills and motivation.	It shows strong empowerment potential but most evidence comes from youth or disease-specific contexts (eg, HIV). Broader transferability has not been tested, limiting PAO's reach.	It bridges education and empowerment but requires validation in chronic and multimorbidity settings for PAO credibility.

Theme	Key studies (authors, year)	Role of biomedical technology	Critical evaluation	Thematic transition
Patient empowerment and co-creation	Vainauskienė and Vaitkienė, 2021 [31]	Real-time feedback dashboards and participatory design platforms encourage collaborative care.	This demonstrates high potential for co-design, but most applications are exploratory or conceptual. Lack of real-world implementation reduces PAO's structural legitimacy.	It completes the feedback loop for PAO but risks tokenism without evidence of genuine patient integration into decision-making.

[a]AI: artificial intelligence.

[b]EHR: electronic health record.

Patient-as-Agent: Emergence of Self-Regulation and Decision-Making

Participants consistently described themselves as "managers" of their health, citing wearables, symptom trackers, and health apps as tools that extend their decision-making processes. They reported scheduling appointments, adjusting lifestyle behaviors, and even questioning clinical advice based on insights derived from digital devices. This reflects a shift toward self-regulation, where patients assume roles once reserved for organizational actors, such as analysts, strategists, and communicators.

> *My smartwatch alerts me when my heart rate spikes, and I've learned to adjust my pace or diet accordingly. It feels like having a personal assistant, but ultimately, I take the final decision.* [Participant 11, age 65 years, rural man, retired government officer]

Digital Health as a Marketing System: Engagement, Nudging, and Feedback Loops

The integration of marketing concepts, particularly segmentation, nudging, and personalization, was evident in how participants responded to app interfaces and notifications. Many acknowledged that gamified elements, personalized reminders, and visual dashboards were crucial for sustaining motivation. However, responses also revealed ethical ambivalence: Although participants appreciated targeted support, they expressed concerns about potential manipulation and the use of data.

> *The app rewards me for reaching my step goals, but I often question how it uses my data. Is it genuinely assisting or trying to sell me something?* [Participant 10, age 30 years, urban woman, renowned corporate figure]

Digital Exclusion and Structural Constraints

Despite enthusiasm for digital tools, disparities were evident. Participants from lower socioeconomic backgrounds or remote locations reported limited access, connectivity challenges, and difficulties navigating complex interfaces. Older participants often lacked digital literacy or felt overwhelmed by "data overload." This highlights the digital divide and the need for inclusive design.

> *The hospital I visited in Delhi told me to use the app, but I don't have Wi-Fi at home. And when I try to use it, it's too confusing. I stop.* [Participant 18, age 55 years, rural woman, school teacher]

Trust and Transparency in AI and Personalization

Participants articulated trust as both a prerequisite and an outcome of digital health interaction. When transparently delivered, personalized care enhanced patients' perception of safety and value. However, algorithmic opacity and inconsistent recommendations undermined confidence. Many desired "explainable AI" and clearer data use policies.

> *If I knew the logic of how it decides what to show me or suggest, I would trust it more. However, it feels like a black box right now.* [Participant 2, age 40 years, urban man, real estate businessman]

Need for Human Touch Amid Digital Expansion

Although digital interfaces were valued for convenience and personalization, patients emphasized the irreplaceable value of human connection. Participants advocated for blended care models where technology augments clinician relationships but does not replace them.

> *I appreciate the app, but I prefer a human to explain serious issues rather than a chatbot.* [Participant 6, age 53 years, suburban woman, homemaker]

Thematic Coding Framework: Operationalizing the PAO Model in Digital Health

To explore how the PAO model is unfolding in real life, we used a 3-step coding process rooted in constructivist-grounded theory. This approach helped us make sense of recurring patterns in what participants shared during interviews.

In the first stage (open coding), we identified specific behaviors and concerns—things like self-tracking habits, reactions to AI-driven personalization, responses to digital nudges, and worries about data privacy and trust (see Table 3). Commercial analogies helped illustrate the PAO model by translating strategies from retail, tech, and service industries to the individual patient level. These comparisons offer fresh perspectives on personalization, engagement, and collaboration, but their nonclinical origins highlight the need for stronger validation within health care settings.

Table . Commercial analogies in the patient as an organization (PAO) model developed by the authors drawing on practices from Amazon, Apple, IKEA, Netflix, and loyalty or supply chain models to illustrate conceptual parallels in the PAO framework.

Commercial practice	Application in the PAO model	What it offers	What it misses
Retail segmentation (eg, Amazon product recommendations)	AI[a]-driven patient segmentation using health and behavioral data to personalize interventions	Tailors care in real time, making treatment more responsive	It risks oversimplifying complex patient needs, and potential for algorithmic bias exists.
Customer loyalty programs (eg, airline frequent flyer, hotel rewards)	Health care CRM[b] platforms that predict adherence and personalize communication	Builds long-term engagement and strengthens patient-provider relationships	Patients are not “customers”—trust in care requires ethical accountability, not just loyalty.
Digital nudging (eg, Netflix autoplay, app notifications)	Health nudges in apps prompting exercise, diet, or medication adherence	Encourages healthy habits and sustained engagement	It may compromise autonomy if patients feel manipulated rather than supported.
Co-creation in services (eg, IKEA design input, open-source platforms)	Participatory health platforms where patients co-design care plans and give feedback	Empowers patients as partners and fosters collaboration	Access barriers and digital literacy gaps may exclude vulnerable populations.
Branding in consumer tech (eg, Apple’s design and trust strategy)	Branding of digital health platforms to foster confidence and ease of use	Reduces anxiety and encourages adoption	Trust in health care must rest on transparency, fairness, and safety, not just design.
Supply chain logistics (eg, just-in-time inventory systems)	Wearables and biosensors providing continuous data for anticipatory care	Prevents crises through early intervention; improves efficiency	It relies on constant connectivity and raises concerns about privacy and governance.

[a]AI: artificial intelligence.

[b]CRM: customer relationship management.

The individual insights shown in Table 3 were then grouped into broader categories during axial coding (Table 4), including themes like digital self-governance, behavioral engagement tools, and trust and ethical friction.

Table . Open coding of everyday experiences with biomedical technology in the patient as an organization (PAO) context, developed by the authors based on qualitative interview data (2025).

Code	Description
Self-tracking	Use of apps or devices to monitor health metrics
Decision autonomy	Making health decisions based on digital feedback
AI[a] personalization	Adjustments based on algorithmic insights
CRM[b]-based reminders	App notifications encouraging health behaviors
Gamification	Points, badges, and visual cues in apps
Behavioral nudging	Subtle prompts guiding patient behavior
Data confusion	Difficulty interpreting or trusting data
Lack of digital access	Limited or no access to the internet or devices
Low digital literacy	Challenges using digital tools due to the skills gap
Patient skepticism	Doubts about data privacy or app motives
Desire for human contact	Preference for in-person over digital interaction
Trust in tech	Confidence in digital tools and recommendations
Transparency concerns	Lack of clarity around data usage and AI processes

[a]AI: artificial intelligence.

[b]CRM: customer relationship management.

Finally, in the selective coding stage, we pulled everything together into 5 core themes (Table 5) that reflect how people are experiencing and adapting to digital health tools: (1) patients as autonomous digital actors, (2) digital health as a behavioral ecosystem, (3) inequities in digital empowerment, (4) negotiating trust and ethical transparency, and (5) blended care as the preferred future. These themes directly address the research’s central question: How is the PAO model operationalized in practice through biomedical technology tools, and what are the ethical, behavioral, and structural implications of this shift?

Table . Axial coding, with code families reflecting the patient as an organization (PAO) model and developed from thematic synthesis of interview data by authors (2025).

Code family	Constituent codes
Digital self-governance	Self-tracking, decision autonomy, AI[a] personalization
Behavioral engagement tools	CRM[b]-based reminders, gamification, behavioral nudging
Structural barriers	Lack of digital access, low digital literacy
Trust and ethical friction	Data confusion, patient skepticism, and transparency concerns
Human-digital synergy	Desire for human contact, trust in tech

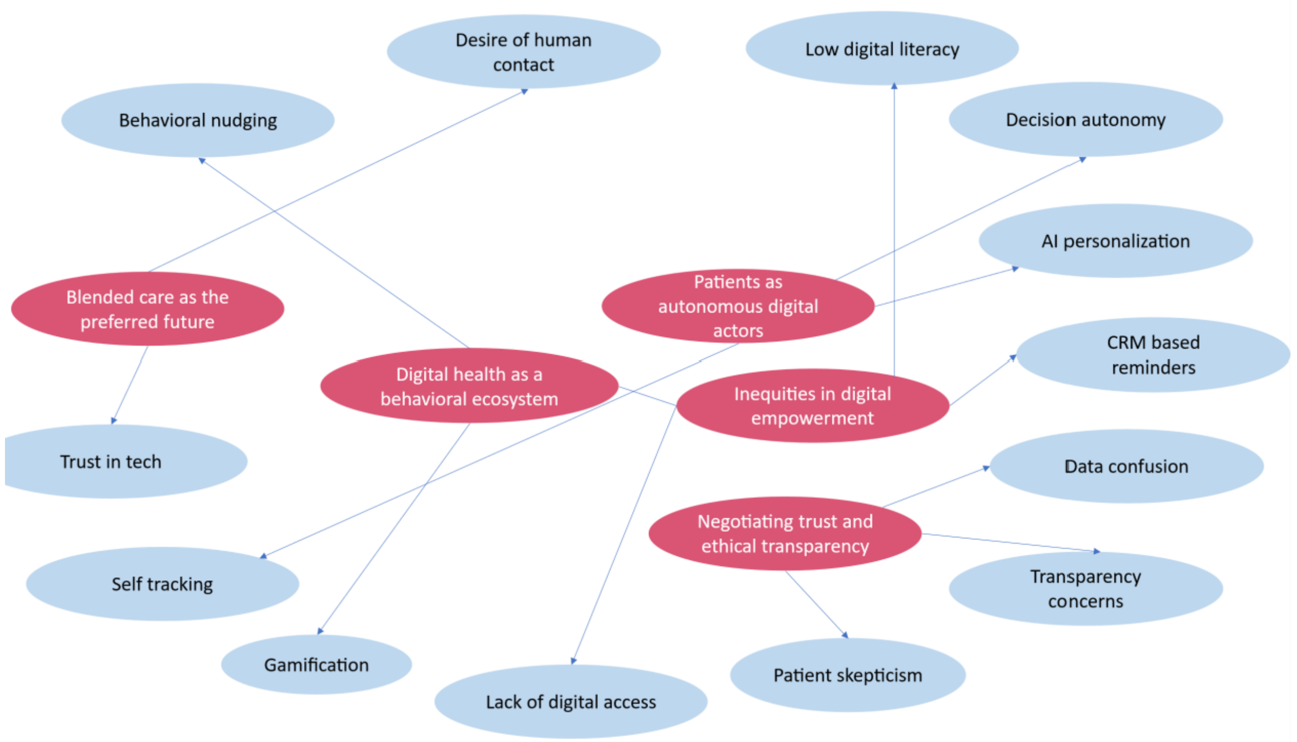
[a]AI: artificial intelligence.

[b]CRM: customer relationship management.

These themes showed how patients are taking a more active role in their care as well as how their experiences are shaped by access, design, trust, and support.

To bring this all to life, Figure 3 maps how individual experiences, like using gamified health apps or struggling with digital literacy, connect to the bigger picture. It visually traces how personal interactions with technology shape and are shaped by the evolving roles patients are playing in today's health care systems.

Figure 3. This code tree, conceptualized by the authors, outlines key patient-centered challenges and themes that shape the digital health experience and documents how individuals interact with emerging technology, the obstacles they face, and the ethical concerns that influence trust and adoption in diverse health care settings. AI: artificial intelligence; CRM: customer relationship management.



This thematic framework grounds the PAO model in real-world voices and helps us better understand how digital tools and health strategies are redefining the patient experience (Table 6). The themes directly address the research's central question: How is the PAO model operationalized in practice through biomedical technology tools, and what are the ethical, behavioral, and structural implications of this shift?

Table . Overarching themes and thematic insights linking patient experiences to the patient as an organization (PAO) framework, as developed from thematic synthesis of interview data by the authors (2025).

Theme	Description	Links to PAO model
Patients as autonomous digital actors	Through digital interfaces, patients demonstrate growing self-regulation and strategic behavior, embodying organizational traits such as monitoring, evaluation, and adaptation.	Aligns with PAO's redefinition of the patient as an active agent in their health journey
Digital health as a behavioral ecosystem	Apps, nudges, CRM[a] reminders, and gamification work in tandem to shape sustained patient engagement. These resemble marketing systems and feedback loops.	Reflects the application of marketing theory (STP[b], CRM) within health care
Inequities in digital empowerment	Access to PAO-enabling technologies is uneven, constrained by structural factors such as socioeconomic status, literacy, and infrastructure.	Reveals a key limitation in PAO implementation across diverse populations
Negotiating trust and ethical transparency	Patients demand clarity around data use and AI[c] decisions. Depending on design and communication, digital systems can facilitate and undermine trust.	Essential for the PAO model to evolve into an ethically grounded framework
Blended care as the preferred future	Despite the convenience of digital technology, human interaction remains essential. Patients seek a hybrid model where technology augments, rather than replaces, the human touch.	Supports a flexible PAO model that integrates human empathy with technological precision

[a]CRM: customer relationship management.

[b]STP: segmentation, targeting, and positioning.

[c]AI: artificial intelligence.

The thematic structure provides a grounded, evidence-based map of how patients are beginning to embody organizational behaviors, where friction points exist, and what conditions are necessary for equitable and sustainable transformation.

Visual Key: Linking Raw Data to Core Themes

In the thematic analysis shown in Figure 3, the red circles represent the high-level themes that emerged during selective coding. These themes, such as *patients as autonomous digital actors* and *blended care as the preferred future*, capture the broader conceptual insights that frame how the PAO model is realized in practice.

In contrast, the blue circles reflect the more granular codes identified during open coding. These codes, such as *self-tracking*, *gamification*, and the *lack of digital access*, are grounded in participants' direct experiences and form the foundation of each theme.

Together, the red and blue circles illustrate how concrete participant narratives (blue) were synthesized into overarching patterns of meaning (red), offering a clear line of sight from real-world observations to theoretical insight. This visual structure is central to understanding the layered complexity of the PAO framework.

Thematic Categories (Axial Coding)

Within the data gathered from the interviews, 5 key themes were evident (Table 5).

The first was the use of digital self-governance. Patients took charge of managing their conditions without assistance by using technology to monitor and track progress.

The second, behavioral engagement tools, covered how reminders, gamification, and nudging led to habit development but also posed questions about user autonomy.

Trust and ethical friction represented the third theme. Attitudes to trust and ethical friction ranged from uncertainty about AI decisions to data use concerns and digital manipulation.

Structural barriers, the fourth theme, included problems such as low digital literacy rates, lack of device and internet access, and poor application usability presented challenges to digital tool use.

In the fifth theme, human-digital synergy involved the use of many valuable digital resources but with an emphasis on augmentation rather than replacement in human interaction.

Integration to Core Themes (Selective Coding)

These factors were integrated to form 5 thematic areas (Table 5) giving insight into the practical application of the PAO model: (1) patients as autonomous digital actors, (2) digital health as a behavioral ecosystem, (3) inequities in digital empowerment, (4) negotiating trust and ethical transparency, and (5) blended care as the preferred future.

These 5 themes formed a complex set of understandings about how people interact with digital health care, not only as technology users but also as strategic and emotionally committed actors. The relationships among these themes are shown in Figure 3, which illustrates how lower-level observations (blue nodes) relate to higher-level organizational themes (red nodes). This level of mapping highlights the richness and depth of analysis in this study and how individual experiences are part of a broader trend in organizational behavior.

Discussion

Principal Findings

Defining the Conceptual Boundaries of the Field

This scoping review advanced the PAO model from a metaphor into a strategic framework rooted in biomedical technology, marketing theory, and digital health practice. Rather than viewing patients as passive care recipients, the literature highlights their role as active, data-aware decision-makers who interact with and shape digital health ecosystems.

The PAO model operates across 3 levels. At the *micro level*, patients manage their health using tools such as wearables, apps, and AI-driven feedback to support self-monitoring and daily decision-making. At the *meso level*, marketing strategies, such as CRM, segmentation, and co-creation, structure how patients engage with health care providers and systems. At the *macro level*, broader issues like policy, regulation, and digital equity influence access, trust, and the overall impact of digital health transformation.

Although the model offers a rich, multilayered view of modern health care, it also presents conceptual challenges, particularly around how individual behaviors translate into system-wide change and how societal structures shape personal experiences. Instead of seeing these tensions as limitations, the review positions them as opportunities to refine the model.

By connecting data flows and decision-making across individual, organizational, and societal levels, the PAO framework has the potential to become a cohesive, ethical, and scalable approach to digital health, one that centers the patient while addressing the realities of technology, governance, and equity.

From Static Segmentation to Dynamic Personalization

Segmentation, targeting, and positioning (STP) have long been foundational marketing strategies. In health care, these have evolved using AI-powered segmentation based on real-time physiological and behavioral data. Studies by Brommels [23] and Minvielle et al [32] described how biomedical technologies enable the continuous reclassification of patients into highly personalized cohorts. However, ethical challenges, particularly those related to algorithmic bias and the risks of overpersonalization, remain underexplored.

Repositioning Behavioral Theories Through Technology

To truly connect biomedical technology with marketing in health care, we need more than comparisons; we need a clear, practical framework. Although strategies such as personalization, CRM, and nudging are often paired with tools like AI and wearables, their true impact on patient behavior is rarely examined.

A better path is to combine behavioral models with real-time tech—using tools that respond to patients' habits, motivations, and health concerns. However, this isn't just a design question; it's an ethical one too. When does a helpful nudge become manipulation?

For the PAO model to be meaningful, it must do more than describe trends. It should explain how these elements work together in everyday care and ensure that technology supports, rather than controls, the patient experience.

Rethinking CRM for a Nonlinear Health Journey

CRM models, which have traditionally been used to foster patient loyalty, now need to adapt to fluid, episodic, and context-driven interactions. Biomedical tools, such as mobile diagnostics and predictive analytics, generate nonlinear touchpoints that challenge traditional CRM funnels [25]. Although CRM provides useful tools for long-term engagement, many studies rely heavily on commercial analogies, neglecting health care–specific factors such as emotional trust, reactions to medical errors, and the maintenance of continuity of care during crises.

Digital Branding and Trust as Ethical Infrastructure

Trust in digital health systems goes beyond user experience; it is built on transparency, data sovereignty, and clarity in algorithmic operations. Mohamed [27] emphasized that branding must now communicate not only usability but also ethical intent. However, most studies treat trust as an outcome rather than a process. There is a need for more in-depth, culturally responsive frameworks that examine how trust develops across various social and institutional contexts.

Adoption Beyond Early Adopters

Innovation adoption is often interpreted through the lens of early adopters, but this perspective overlooks the systemic factors that lead certain populations to resist digital health tools [33]. These include limited infrastructure, historical mistrust, and mismatched cultural values. The existing literature lacks a pluralistic adoption model that reflects the social, historical, and geographic nuances that influence adoption behavior.

The Ethical Ambiguity of Nudging

The conceptual model often presents digital tools, such as AI platforms, wearables, and CRM systems, as if they will inherently empower patients, foster engagement, and enhance autonomy. Although these technologies hold great promise, including personalized care, deeper engagement, and patient co-creation, this view risks overlooking important challenges, including bias, inequality, and subtle forms of coercion (Table 7). Patients may experience technology fatigue, struggle with data misinterpretation, or face structural inequities that limit access and benefits. At the same time, behavioral economics and nudging strategies, though effective in shaping healthier choices, blur the line between ethical persuasion and unintended coercion. In high-stakes health contexts, design elements intended to encourage positive behaviors can inadvertently pressure or manipulate patients, undermining trust. However, few studies critically distinguish between user-centered design that supports autonomy and subtle mechanisms of control that erode it. For the PAO model to develop into a strong framework, it must include a more balanced view—one that recognizes both the potential of digital tools and the ethical and structural risks they pose. This balance will be key for creating frameworks that empower patients without sacrificing ethical safeguards, equitable access, agency, or trust.

Table . Empowering potential and risks of digital tools in the patient as an organization (PAO) model, as developed by the authors based on concepts from digital health, behavioral economics, and health care marketing literature.

Digital tools or strategies	Potential benefits	Associated risks
AI[a] platforms and predictive analytics	Deliver personalized recommendations, enable anticipatory care, and support clinical decision-making	Risk of algorithmic bias, misinterpretation of complex data, and over-reliance on automated outputs
Wearables and biosensors	Provide real-time monitoring, support self-management, and allow early detection of health concerns	Can lead to technology fatigue, data overload, and unequal access due to cost or connectivity
CRM[b] systems in health care	Strengthen patient-provider relationships, tailor communication, and encourage proactive engagement	May reduce patients to "customers," raise privacy concerns, or foster dependency on system-generated prompts
Digital nudging and behavioral economics	Encourage healthier behaviors, improve adherence, and reinforce positive routines	Raise ethical concerns about manipulation, risk of coercion in high-stakes decisions, and potential erosion of autonomy
Participatory platforms and co-creation	Promote shared decision-making, build trust, and empower patients as partners in care	Digital literacy gaps, exclusion of marginalized groups, and uneven levels of participation

[a]AI: artificial intelligence.

[b]CRM: customer relationship management.

From Information to Empowerment: Revisiting the Information-Motivation-Behavioral Skills Model

The information-motivation-behavioral skills (IMB) model is evolving from an educational tool into an empowerment framework, supported by AI-driven guidance systems, decision aids, and adaptive learning tools [30]. However, questions persist: Who determines which information is relevant? How is motivation culturally constructed?

There is a need for more critical, reflexive research that challenges normative assumptions about information delivery and authority.

Feedback and Co-Creation: From Input to Shared Power

Co-creation platforms and feedback mechanisms are central to participatory health care, but many current models stop at surface-level input. Without mechanisms for integrating patient feedback into system design and decision-making, participation risks becoming symbolic [31]. Few studies distinguish between procedural engagement and substantive influence, leaving a gap in conceptualizing truly collaborative health systems.

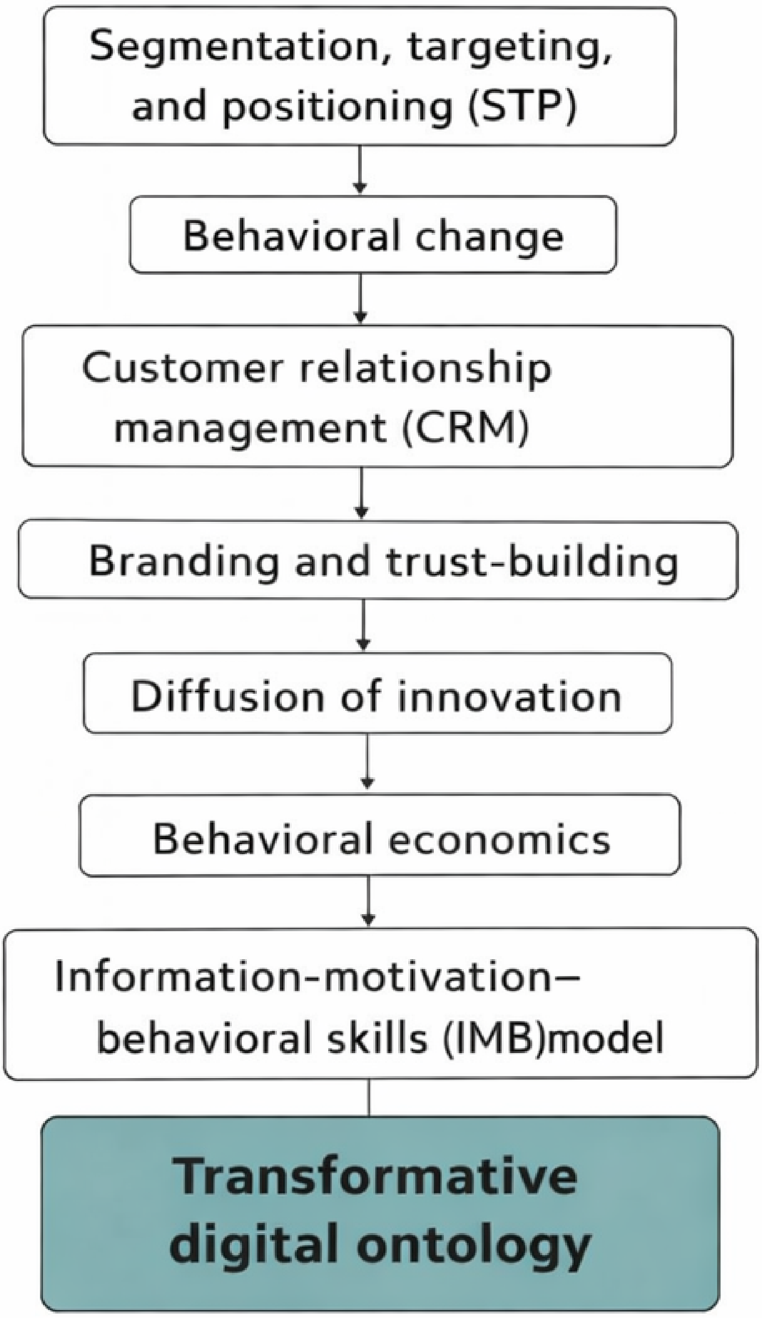
Together, these insights suggest a paradigmatic shift in how we define and interact with the digital patient. The PAO model, when fully realized, reframes the patient as a co-manager, system shaper, and strategic partner in care. This review clarifies the field's theoretical boundaries and proposes an interdisciplinary foundation integrating biomedical engineering, ethical marketing, behavioral science, and patient co-agency. As the digital health landscape continues to evolve, this lens provides researchers, designers, and policymakers with a critical roadmap for developing ethical, inclusive, and technologically responsive systems.

Toward a Conceptual Contribution

The PAO model is often described as if individuals could fully take on the roles of structured organizations, overseeing strategy, governance, and operations. Although this metaphor effectively represents patient empowerment, it risks oversimplifying complex realities and may overestimate what patients can do. Unlike formal organizations, patients usually lack dedicated resources, hierarchical leadership, and institutional authority. Without a clearer definition of what "organizational behavior" means at the individual level, the PAO risks becoming more of a rhetorical device than a practical framework.

This synthesis offers an opportunity to rethink the PAO model, not as a collection of marketing ideas but as a transformative digital framework. As shown in Figure 4, this new approach builds on a step-by-step integration of segmentation, trust-building, behavioral economics, and co-creation strategies, culminating in the development of a transformative digital ontology. Together, these mechanisms change the patient's role from a passive recipient of care to an active, data-informed participant in health care ecosystems. Biomedical technologies are central to this shift: By combining data, autonomy, and organizational logic, they redefine what it means to be a "patient," creating new forms of digital identity that are both empowered and connected [34].

Figure 4. Sequential integration of marketing theories into the patient as an organization (PAO) framework, working toward a conceptual contribution of the patient as a strategic, participatory actor within a dynamic digital health ecosystem and based on the authors' thematic synthesis of literature included in the scoping literature review.



It is important to recognize that the shift toward the PAO model is not equally accessible to all. How patients engage with digital tools is shaped by power dynamics, design decisions, access gaps, and governance structures. Acknowledging these realities does not weaken the PAO concept—it strengthens it by promoting a more inclusive and critical approach to digital health. To advance the field, research must move beyond idealized visions and explore how these dynamics play out in practice. A critical digital health perspective is essential: one that asks what technologies do, what kinds of patient roles they create, and who truly benefits. Rather than undermining the PAO model, this approach ensures it evolves ethically and equitably, with patient trust at its core. By embracing both its potential and its limitations, the PAO model can become a meaningful framework to guiding the future of patient-centered digital health care.

Table 2 highlights the potential of biomedical technologies and marketing frameworks to bring the PAO model into practice but also reveals gaps in the evidence.

Inference and Interpretive Insight

These findings demonstrate that the PAO model is somewhat implemented in practice, especially among digitally engaged patients. However, its application varies and is contingent on access, literacy, trust, and ethical clarity. Although digital tools are starting to promote organizational behaviors, such as self-monitoring, treated responses based on segmentation, and adaptations driven by feedback, they are primarily effective within privileged structures [35].

Participants appreciate hyper-personalization but seek equity, transparency, and human connection. These results underscore the need for interdisciplinary collaboration to ensure that digital health systems are technologically advanced, ethically sound, and socially inclusive.

Repositioning the Patient: Conceptual and Thematic Foundations of the PAO Model

This analysis reframed the PAO model by combining insights from biomedical innovation and marketing theory, viewing patients as active, data-driven participants in digital health systems. Traditionally, patient engagement focused on clinical outcomes or behavior change. The PAO model expands on this by positioning patients as strategic actors supported by tools such as wearables, mHealth apps, AI, and predictive analytics.

The scoping review identified 8 recurring themes where marketing concepts such as STP, CRM, behavioral economics, and co-creation intersect with biomedical technologies to create adaptive, personalized care ecosystems. Patients increasingly take on organizational roles including *strategy*, using data to set health priorities; *governance*, managing consent, trust, and accountability; and *operations*, coordinating daily care with digital tools.

Marketing analogies help clarify this shift. AI segmentation mirrors retail targeting but is used here for real-time health decisions. CRM becomes a patient-care platform, while nudging and digital branding influence health behaviors, much like tech firms shape user choices.

Although these analogies make the PAO model relatable, they rely heavily on conceptual models and pilot studies, lacking the clinical rigor expected in health care. The model's strength lies in making new patient roles visible; its weakness lies in limited empirical support.

Ultimately, the PAO model moves beyond metaphor, presenting patients as informed, autonomous agents in digitally enabled care. However, real-world challenges, like power imbalances, access barriers, and ethical risks, remain. For the model to mature, it must blend the creativity of marketing insights with the credibility of clinical and behavioral evidence.

Key Research Gaps in the PAO Model and Marketing-Driven Digital Health With Biomedical Integration

Limited Empirical Validation of the PAO Model

Although the PAO model offers strong conceptual foundations, there is a notable lack of empirical studies demonstrating how patients engage in organizational-like behaviors using digital health technologies. The existing literature tends to focus on potential rather than observed behaviors [36]. To substantiate the PAO framework, there is an urgent need for longitudinal, ethnographic, and practice-based research that captures patient engagement across diverse cultural and clinical contexts.

Inconsistent Application of Marketing Theories in Health Care

Despite the established value of marketing frameworks such as STP, CRM, and behavioral marketing in other industries, their integration in health care remains fragmented. The literature review revealed a lack of standardization in the application of these theories to influence patient engagement or health outcomes. There is a research gap in developing an evidence-based framework that operationalizes these theories via biomedical technologies and ties them to measurable behavioral or clinical outcomes [37].

Underexplored Intersection of Health Equity and Digital Access

The review highlighted persistent health disparities rooted in socioeconomic status, geography, and digital literacy [38]. Although these structural barriers are well-documented individually, few studies explore how they intersect with the PAO framework or shape patient access to digital health tools. Future research should prioritize equity-focused design and examine how inclusive strategies can effectively close the digital divide in practice, rather than just in theory.

Neglected Ethical and Regulatory Dimensions of Digital Health Marketing

Ethical concerns, such as informed consent, data transparency, algorithmic bias, and digital nudging, are acknowledged across multiple sources [39]. However, in-depth theoretical engagement remains sparse. Most studies briefly address these issues without providing analytical frameworks or policy recommendations. This gap underscores the need for comprehensive ethical models that can guide the development and deployment of AI-driven marketing in health care settings.

Insufficient Understanding of AI's Impact on Trust and Autonomy

Although AI offers significant potential for care personalization, the implications for patient trust, autonomy, and long-term relationships with health care providers are poorly understood [40]. The review revealed a lack of studies examining how predictive analytics influence user experience, clinical decision-making, or interpersonal dynamics in digital care environments. Responsible AI implementation requires empirical research that evaluates these relational dimensions.

Lack of Empirical Work on Nudging and Behavioral Economics

Behavioral economics and digital nudging are frequently cited as promising tools for influencing health behaviors [41]. However, the literature offers few empirical studies on their effectiveness in complex or high-stakes medical decisions. Experimental designs and real-world field studies are necessary to investigate how interface design, default options, and incentive framing influence behavior, particularly when ethical boundaries are being tested.

The Need for an Integrated, Evidence-Based Framework

Perhaps the most critical gap identified was the lack of a comprehensive PAO framework that cohesively combines marketing theory, behavioral psychology, and biomedical innovation. Current research is scattered across disciplinary silos and lacks a unified theory that can capture the complexity of digitally enabled, ethically nuanced patient engagement [42]. Filling these gaps requires a cross-disciplinary approach that is both essential and collaborative. Researchers must go beyond conceptual enthusiasm and move toward empirical validation, creating inclusive, ethically sound systems that empower patients without compromising trust or equity [43]. Only then can the PAO model develop into a practical, evidence-based foundation for personalized and participatory digital health care.

This study revealed a powerful shift underway in health care: Patients are no longer just following care plans; they're actively shaping them. With the support of digital tools like wearables, mobile apps, and AI-driven platforms, many are setting health goals, tracking their progress, and making informed decisions. These behaviors closely resemble how organizations operate [44].

The scoping review showed that technologies are increasingly infused with marketing strategies like personalization, segmentation, CRM, and digital nudging. In interviews, these concepts came to life. Patients were not just using apps to manage symptoms; they were navigating complex decisions, often independently, guided by digital feedback.

However, this transformation isn't universal. Although some patients found empowerment, others faced barriers, limited digital access, low tech literacy, or distrust in AI. Alongside enthusiasm, participants expressed concerns about losing human touch, data misuse, or feeling subtly manipulated by digital nudges [45].

These tensions reveal a deeper truth: The future of health care isn't just about technology; it's about how that technology is designed, delivered, and experienced. Patients want digital tools that support, not replace, human care. Many envisioned a *blended model*—one where empathy and innovation go hand in hand [46].

Ultimately, the PAO model is no longer just a metaphor. It's emerging in real life but unevenly. To realize its full potential, future systems must be co-created with patients, grounded in trust, and built for equity not just efficiency.

This research shows that a quiet transformation is happening in medicine. Patients are no longer just recipients; they are now active participants in managing their health experiences. Through technologies such as wearables, apps, and AI platforms, patients are taking control and making real-time decisions tailored to their needs [47]. As a result, these patients are assuming responsibilities that are increasingly like self-management and self-organization.

However, this shift involves more than just technology. It also depends on how people feel, how they trust, whether they think they can maintain control, and whether they believe the systems support them. Although some tools, such as digital nudging and CRM reminders, help patients stay on track, others can make them feel anxious. Concerns about data privacy, algorithmic data misuse, and the loss of the human touch are common.

What is clear is that this transformation has great potential if approached thoughtfully. Empowerment must be balanced with protection. For the PAO model to guide future care, it needs to reflect what technology can do and what people want to achieve.

Comparison With Prior Work

Prior work in digital health has practically proven escalating patient engagement with the assistance of vehicles such as wearables, mHealth apps, and AI-driven platforms. Ahadzadeh et al [24], for example, examined how the integration between the health belief model and the technology acceptance model dictated patient behavior in digital environments. In contrast, Rongkavilit et al [30] applied the IMB model to study medication behavior in teenagers infected with HIV/AIDS. These indicate the potential of digital tools to enhance motivation, information access, and behavioral skills [30]. These, however, primarily depict patients' empowerment resulting from educational exposure or risk awareness, with little attention to patients as agency-rich actors with the potential for system-level agency.

Reframing Empowerment Within the PAO Model

In addition, this research contributes to the expanding body of digital health literature, as it recasts this concept of empowerment through the PAO framework. This indicates that, as part of this paradigm, patients are not just receiving and using digital health technology but are strategic and savvy about their own data, as would be expected of organizations, not humans.

Both personalization and nudging have been touted as means of improving engagement, yet, as we explored through interviews, a more nuanced picture is emerging. Concerns about lack of transparency and accountability, as well as manipulation, have been voiced by patients, emphasizing that patient empowerment is more than mere behavior—it is ethical as well [48].

Key Contributions

Adding Patient Organizational and Strategic Roles as an Expansion of the Idea of Empowerment

Behavior model integration (eg, HBM, IMB, or CRM) into a multilevel, ethics-focused PAO framework addresses structural inequality and moral controversies, especially for marginalized user communities, by providing practical validation through patient experiences that reveal both the potential and the vulnerabilities of technology.

Structural Inequities

There is inequality in access to digital care. Some participants, especially those lacking sufficient digital literacy and computer access, experienced identifiable barriers described under the "structural barriers" code (Table 5), which is encapsulated under inequities in digital empowerment (Table 6). Such barriers, as identified in Figure 3, lie at the intersections of trust, accessibility, and ability, emphasizing that care systems should be built for all and not just for digital sophisticates.

Trust, Ethics, and Digital Engagement

By contrast, trust was a motivator and an inhibitor of engagement. Doubts and reservations about data usage or about algorithms and nudging, coded under trust and ethical friction (Table 5), often revolve around ethical issues that have been aggregated into the theme of negotiating trust and ethical transparency (Table 6), emphasizing that, although medical systems need to be efficient, they should be transparent and eminently explainable and that patient autonomy must be respected.

Finding Meaning and Connection

Despite this momentum in confidence in digital technology, one key takeaway was evident—the need for emotional connection remained. Apps and AI technology might be incredibly convenient and informative, but in no way could these technologies provide emotional understanding, particularly with emotional experiences and more complex medical choices. The best possible solution continued to be that found within the process of blended care (see "blended care as the preferred future" in Figure 3), where digital technology interacted with emotional connection and interpersonal trust that was found to be absolutely critical to the process of care in the digital age.

Strengths and Weaknesses

A key strength of this study was its mixed methods approach, which combines a broad literature review with rich qualitative insights. The use of grounded theory enabled the emergence of nuanced, real-world themes that reflect the diversity of patient experiences.

However, the study was limited by a relatively small and context-specific sample in the qualitative phase. Broader validation across diverse populations and health care systems will be needed to assess generalizability. Additionally, the scoping review was conceptual in nature and did not include a formal risk-of-bias assessment.

Contributions to Theory and Practice

This study makes several contributions to the evolving digital health literature. It *redefines patient empowerment* as a multilevel, ethically grounded process that includes digital skills, trust-building, and system-level support. It integrates *behavioral models* (eg, health belief model, CRM) with real-time technologies to understand how patients engage with care dynamically. It adds patient perspectives on *equity, ethics, and lived experience*, which are often missing from technologically focused research. It validates the PAO model through empirical data, showing how patients actively manage care but also struggle with ethical ambiguity and structural barriers.

Future Directions

Future studies should consider equity-oriented design and co-creation as well as research into trust, behavior, and adoption over time or other studies on ethical metrics for nudging, personalization, and AI adoption and use in patient care and outcomes today. It is imperative that, in the future, a PAO model be one of innovation, grounded in patients' experiences, anxieties, and values.

To ensure that the PAO model becomes an empirical reality and shifts from theoretical to practical application, the following should be prioritized in subsequent research: (1) equity-informed design: design technologies to be accessible to everyone; (2) co-creation with patients: engage patients in the design process, evaluation, and governance of digital health applications; (3) transparent and ethical infrastructures: develop consent dashboards and explainable AI; and (4) longitudinal study research: study digital behavior trust processes over time.

Conclusion

This paper examined another important change that has come about in the health care sector. Patients are no longer passive receivers of health care but are actively participating in it with the use of technology such as wearables, health apps, and AI platforms. Technology is giving people the capacity to take charge of their health care choices [49].

We conducted our study through the review of existing literature and in-depth interviews to explore how the application of marketing approaches, such as personalization, nudging, and CRM, is integrated with digital health platforms. The PAO conceptual framework provides valuable insights to understand this transformation and recognizes that patients are "consumers" no more but act like strategic actors who use data to manage health just like any other organizational entity.

What our research tells us is that, for some patients, particularly those with robust digital connectivity and savvy, this scenario has already begun to come to fruition. Many of the patients surveyed consider themselves to be planners or "health managers," utilizing digital feed-forward and tracking applications to shape decisions. However, such technologies are certainly not ubiquitous. Members from underserved groups presented authentic challenges to digital health adoption, such as connectivity difficulties and unfamiliarity with AI applications.

Additionally, there are other ethical considerations that must be considered. Although data personalization and nudging can be beneficial to patient health outcomes, there are worries about transparency or manipulation and autonomy if patients do not fully understand algorithmic influences [50].

Despite these challenges, the PAO model provides significant insight into what is happening to the role of the patient. Rather than viewing it as something that is complete and finished—something that needs to be applied—the best way to look at it is to realize that it can be seen as more of a thought process about the future of health care that must allow for adjustment and change.

In the end, this model encourages us to reconsider the role of patients—be it within health care as recipients of care but also as active contributors to the creation thereof. With such caution and care taken in its development, the PAO approach might lead to a more participatory and trustable future in health. Only then can we ensure that digital health is not just innovative but also inclusive, trustworthy, and deeply human.

Acknowledgments

The figures (except Figure 4) in this manuscript were designed by Anusha Das Gupta (Cornell University), whose visual contributions significantly enhanced the presentation of the study’s conceptual and analytical framework. Generative artificial intelligence (AI) tools were used in a limited and supportive capacity during manuscript preparation. Figure 4 was created with the assistance of SORA to help visualize the patient as an organization (PAO) model, and Grammarly was used to improve grammar and readability in draft versions. Importantly, no AI tools were involved in data collection, analysis, or interpretation. All authors carefully reviewed and approved the final content and take full responsibility for its accuracy and integrity.

Funding

No funding was received from any sources.

Data Availability

The datasets generated during and/or analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: ADG
Data curation: YY
Formal analysis: ADG
Investigation: YY
Methodology: ADG
Validation: YY
Visualization: ADG
Writing – original draft: ADG
Writing – review & editing: YY

Conflicts of Interest

None declared.

Checklist 1
PRISMA-ScR checklist.
[DOCX File, 18 KB - biomedeng_v11i1e77115_app1.docx]

Checklist 2
COREQ checklist.
[DOCX File, 16 KB - biomedeng_v11i1e77115_app2.docx]

References

1. WHO global strategy on people-centred and integrated health services: interim report. World Health Organization. 2015. URL: https://apps.who.int/iris/handle/10665/155002 [accessed 2025-12-07]
2. Al Kuwaiti A, Nazer K, Al-Reedy A, et al. A review of the role of artificial intelligence in healthcare. J Pers Med 2023 Jun 5;13(6):951. [doi: 10.3390/jpm13060951] [Medline: 37373940]
3. National Academies of Sciences, Engineering, and Medicine. Implementing High-Quality Primary Care: Rebuilding the Foundation of Health Care: National Academies Press; 2021. [doi: 10.17226/25983]
4. Tahri Sqalli M, Al-Thani D. On how chronic conditions affect the patient-AI interaction: a literature review. Healthcare (Basel) 2020 Sep 1;8(3):313. [doi: 10.3390/healthcare8030313] [Medline: 32883036]
5. Mennella C, Maniscalco U, De Pietro G, Esposito M. Ethical and regulatory challenges of AI technologies in healthcare: a narrative review. Heliyon 2024 Feb 29;10(4):e26297. [doi: 10.1016/j.heliyon.2024.e26297] [Medline: 38384518]
6. Haleem A, Javaid M, Singh RP, Suman R. Telemedicine for healthcare: capabilities, features, barriers, and applications. Sens Int 2021;2:100117. [doi: 10.1016/j.sintl.2021.100117] [Medline: 34806053]
7. Betz UAK, Arora L, Assal RA, et al. Game changers in science and technology - now and beyond. Technol Forecast Soc Change 2023 Aug;193:122588. [doi: 10.1016/j.techfore.2023.122588]
8. Fusco F, Marsilio M, Guglielmetti C. Co-creation in healthcare: framing the outcomes and their determinants. Journal of Service Management 2023 Dec 18;34(6):1-26. [doi: 10.1108/JOSM-06-2021-0212]
9. Kruk ME, Gage AD, Arsenault C, et al. High-quality health systems in the Sustainable Development Goals era: time for a revolution. Lancet Glob Health 2018 Nov;6(11):e1196-e1252. [doi: 10.1016/S2214-109X(18)30386-3] [Medline: 30196093]

10. Shajari S, Kuruvinashetti K, Komeili A, Sundararaj U. The emergence of AI-based wearable sensors for digital health technology: a review. Sensors (Basel) 2023 Nov 29;23(23):1-36. [doi: 10.3390/s23239498] [Medline: 38067871]
11. Leao DLL, Cremers HP, van Veghel D, Pavlova M, Groot W. The impact of value-based payment models for networks of care and transmural care: a systematic literature review. Appl Health Econ Health Policy 2023 May;21(3):441-466. [doi: 10.1007/s40258-023-00790-z] [Medline: 36723777]
12. Unanah OV, Mbanugo OJ. Integration of AI into CRM for effective U.S. healthcare and pharmaceutical marketing. World J Adv Res Rev 2025;25(2):609-630. [doi: 10.30574/wjarr.2025.25.2.0396]
13. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
14. Junaid SB, Imam AA, Balogun AO, et al. Recent advancements in emerging technologies for healthcare management systems: a survey. Healthcare (Basel) 2022 Oct 3;10(10):1940. [doi: 10.3390/healthcare10101940] [Medline: 36292387]
15. Shahzad MF, Xu S, Javed I. ChatGPT awareness, acceptance, and adoption in higher education: the role of trust as a cornerstone. Int J Educ Technol High Educ 2024;21(1):1. [doi: 10.1186/s41239-024-00478-x]
16. Gehman J, Glaser VL, Eisenhardt KM, Gioia D, Langley A, Corley KG. Finding theory–method fit: a comparison of three qualitative approaches to theory building. Journal of Management Inquiry 2018 Jul;27(3):284-300. [doi: 10.1177/1056492617706029]
17. Morse JM. Determining sample size. Qual Health Res 2000 Jan;10(1):3-5. [doi: 10.1177/104973200129118183]
18. Frequently asked questions (faqs) on new drugs and clinical trials. Central Drugs Standard Control Organization Directorate General of Health Services Ministry of Health and Family Welfare Government of India. URL: https://cdsco.gov.in/opencms/resources/UploadCDSCOWeb/2018/UploadPublic_NoticesFiles/faqnd.pdf [accessed 2025-12-11]
19. No DCGI approval required for clinical trials for academic purposes. PSA Legal Counsellors. URL: https://www.psalegal.com/wp-content/uploads/2021/12/No-DCGI-approval-required-for-clinical-trials-for-academic-purposes.pdf [accessed 2025-12-11]
20. CDSA. Clinical trials toolkit. Translational Health Science and Technology Institute. URL: https://cdsatoolkit.thsti.in [accessed 2025-12-11]
21. Shaw L, Jazayeri D, Kiegaldie D, Morris M. Virtual communities of practice to improve clinical outcomes in healthcare: protocol for a 10-year scoping review. BMJ Open 2021 Jul 26;11(7):e046998. [doi: 10.1136/bmjopen-2020-046998] [Medline: 34312199]
22. Christou PA. How to use thematic analysis in qualitative research. Journal of Qualitative Research in Tourism 2023;3(2):79-95. [doi: 10.4337/jqrt.2023.0006]
23. Brommels M. Patient segmentation: adjust the production logic to the medical knowledge applied and the patient's ability to self-manage-a discussion paper. Front Public Health 2020;8:195. [doi: 10.3389/fpubh.2020.00195] [Medline: 32537449]
24. Ahadzadeh AS, Pahlevan Sharif S, Ong FS, Khong KW. Integrating health belief model and technology acceptance model: an investigation of health-related internet use. J Med Internet Res 2015 Feb 19;17(2):e45. [doi: 10.2196/jmir.3564] [Medline: 25700481]
25. Evans WD. What social marketing can do for you. BMJ 2006 Aug 5;333(7562):299. [doi: 10.1136/bmj.333.7562.299]
26. Mohiuddin AK. Patient relationship management: a CRM approach to patient care. Int J Heal Clin Res 2019;2(2):1-11 [FREE Full text]
27. Mohamed NSP. Branded healthcare environment: building trust among patients. ResearchGate. 2022. URL: https://www.researchgate.net/publication/358467936_Branded_Healthcare_Environment_Building_Trust_among_Patients [accessed 2025-12-07]
28. Dearing JW, Cox JG. Diffusion of innovations theory, principles, and practice. Health Aff (Millwood) 2018 Feb;37(2):183-190. [doi: 10.1377/hlthaff.2017.1104] [Medline: 29401011]
29. Auf H, Dagman J, Renström S, Chaplin J. Gamification and nudging techniques for improving user engagement in mental health and well-being apps. Proc Des Soc 2021 Aug;1:1647-1656. [doi: 10.1017/pds.2021.426]
30. Rongkavilit C, Naar-King S, Kaljee LM, et al. Applying the information-motivation-behavioral skills model in medication adherence among Thai youth living with HIV: a qualitative study. AIDS Patient Care STDS 2010 Dec;24(12):787-794. [doi: 10.1089/apc.2010.0069] [Medline: 21091238]
31. Vainauskienė V, Vaitkienė R. Enablers of patient knowledge empowerment for self-management of chronic disease: an integrative review. Int J Environ Res Public Health 2021 Feb 24;18(5):1-24. [doi: 10.3390/ijerph18052247] [Medline: 33668329]
32. Minvielle E, Waelli M, Sicotte C, Kimberly JR. Managing customization in health care: a framework derived from the services sector literature. Health Policy 2014 Aug;117(2):216-227. [doi: 10.1016/j.healthpol.2014.04.005] [Medline: 24837516]
33. Zhang X, Yu P, Yan J, Ton A M Spil I. Using diffusion of innovation theory to understand the factors impacting patient acceptance and use of consumer e-health innovations: a case study in a primary care clinic. BMC Health Serv Res 2015 Feb 21;15(1):71. [doi: 10.1186/s12913-015-0726-2] [Medline: 25885110]
34. Das Gupta A. Conceptualizing patient as an organization with the adoption of digital health. Biomed Eng Comput Biol 2024;15:11795972241277292. [doi: 10.1177/11795972241277292] [Medline: 39324148]

35. Cheng ZM, Bonetti F, de Regt A, Ribeiro JL, Plangger K. Principles of responsible digital implementation: developing operational business resilience to reduce resistance to digital innovations. Organ Dyn 2024 Apr;53(2):101043. [doi: 10.1016/j.orgdyn.2024.101043]
36. Lawson J, Cabili MN, Kerry G, et al. The Data Use Ontology to streamline responsible access to human biomedical datasets. Cell Genom 2021 Nov 10;1(2). [doi: 10.1016/j.xgen.2021.100028] [Medline: 34820659]
37. Liao Y, Thompson C, Peterson S, Mandrola J, Beg MS. The future of wearable technologies and remote monitoring in health care. Am Soc Clin Oncol Educ Book 2019 Jan;39(39):115-121. [doi: 10.1200/EDBK_238919] [Medline: 31099626]
38. Saeed SA, Masters RMR. Disparities in health care and the digital divide. Curr Psychiatry Rep 2021 Jul 23;23(9):61. [doi: 10.1007/s11920-021-01274-4] [Medline: 34297202]
39. Tsamados A, Aggarwal N, Cowls J, et al. The ethics of algorithms: key problems and solutions. AI & Soc 2022 Mar;37(1):215-230. [doi: 10.1007/s00146-021-01154-8]
40. Sauerbrei A, Kerasidou A, Lucivero F, Hallowell N. The impact of artificial intelligence on the person-centred, doctor-patient relationship: some problems and solutions. BMC Med Inform Decis Mak 2023 Apr 20;23(1):73. [doi: 10.1186/s12911-023-02162-y] [Medline: 37081503]
41. Matjasko JL, Cawley JH, Baker-Goering MM, Yokum DV. Applying behavioral economics to public health policy: illustrative examples and promising directions. Am J Prev Med 2016 May;50(5 Suppl 1):S13-S19. [doi: 10.1016/j.amepre.2016.02.007] [Medline: 27102853]
42. Mumtaz H, Riaz MH, Wajid H, et al. Current challenges and potential solutions to the use of digital health technologies in evidence generation: a narrative review. Front Digit Health 2023;5:1203945. [doi: 10.3389/fdgth.2023.1203945] [Medline: 37840685]
43. Adegbesan A, Akingbola A, Ojo O, et al. Ethical challenges in the integration of artificial intelligence in palliative care. Journal of Medicine, Surgery, and Public Health 2024 Dec;4:100158. [doi: 10.1016/j.glmedi.2024.100158]
44. Tong HL, Maher C, Parker K, et al. The use of mobile apps and fitness trackers to promote healthy behaviors during COVID-19: a cross-sectional survey. PLOS Digit Health 2022 Aug;1(8):e0000087. [doi: 10.1371/journal.pdig.0000087] [Medline: 36812578]
45. Shabankareh M, Ranjbaran A, Lim WM, Daronkola HK, Nazarian A. Artificial intelligence on social media and revisit behavior: evidence from the hotel industry. International Journal of Advertising 2025:1-40. [doi: 10.1080/02650487.2025.2536361]
46. Giansanti D. The future of healthcare is digital: unlocking the potential of mobile health and e-Health solutions. Healthcare (Basel) 2025 Apr 3;13(7):802. [doi: 10.3390/healthcare13070802] [Medline: 40218099]
47. Maleki Varnosfaderani S, Forouzanfar M. The role of AI in hospitals and clinics: transforming healthcare in the 21st century. Bioengineering (Basel) 2024 Mar 29;11(4):337. [doi: 10.3390/bioengineering11040337] [Medline: 38671759]
48. National Academy of Medicine. Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril: National Academies Press; 2023. [doi: 10.17226/27111]
49. Kang HS, Exworthy M. Wearing the future-wearables to empower users to take greater responsibility for their health and care: scoping review. JMIR Mhealth Uhealth 2022 Jul 13;10(7):e35684. [doi: 10.2196/35684] [Medline: 35830222]
50. Beermann V. The three waves of digital nudging: concepts, mechanisms, and ethical implications for information systems. Presented at: America Conference on Information Systems (AMCIS); Aug 14-16, 2025 URL: https://www.researchgate.net/publication/394517287_The_Three_Waves_of_Digital_Nudging_The_Three_Waves_of_Digital_Nudging_Concepts_Mechanisms_and_Ethical_Implications_for_Information_Systems [accessed 2025-12-16]

Abbreviations

AI: artificial intelligence
CRM: customer relationship management
IMB: information-motivation-behavioral skills
mHealth: mobile health
PAO: patient as an organization
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
RQ: research question
STP: segmentation, targeting, and positioning
WHO: World Health Organization

Edited by J Sarvestan; submitted 07.May.2025; peer-reviewed by M Gasmi, RSG Mahmoud, YW Shau; revised version received 11.Nov.2025; accepted 12.Nov.2025; published 06.Jan.2026.

Please cite as:
Das Gupta A, Yadav Y
Conceptual Model for the Integration of Marketing Strategies and Biomedical Innovation in Patient-Centered Care: Mixed Methods Study
JMIR Biomed Eng 2026;11:e77115
URL: https://biomedeng.jmir.org/2026/1/e77115
doi:10.2196/77115

Increasing Large Language Model Accuracy for Care-Seeking Advice Using Prompts Reflecting Human Reasoning Strategies in the Real World: Validation Study

Marvin Kopka, BSc, MSc, MPH, Dr rer medic, PhD; Markus A Feufel, Dipl-Ing (FH), MSc, PhD

Division of Ergonomics, Department of Psychology & Ergonomics (IPA), Technische Universität Berlin, Straße des 17. Juni 135, Berlin, State of Berlin, Germany

Corresponding Author:
Marvin Kopka, BSc, MSc, MPH, Dr rer medic, PhD
Division of Ergonomics, Department of Psychology & Ergonomics (IPA), Technische Universität Berlin, Straße des 17. Juni 135, Berlin, State of Berlin, Germany

Abstract

Background: Current prompting techniques for large language models (LLMs), such as ChatGPT, mainly focus on well-structured, low-uncertainty problems; yet, many real-world tasks (eg, care-seeking decisions) are ill-defined and involve high uncertainty. Naturalistic decision-making (NDM) specifically analyzes how humans make accurate decisions in such settings, but NDM concepts have not yet been applied to LLM prompt engineering.

Objective: This study aimed to determine whether prompting strategies inspired by NDM (specifically based on recognition-primed decision-making and the data-frame theory) could improve LLM performance in a real-world, high-uncertainty task, such as making care-seeking decisions.

Methods: We evaluated 10 ChatGPT models (GPT-4o, GPT-4.1, GPT-4.1 mini, o3, o4 mini, o4 mini high, GPT-5.1 Instant, GPT-5.1 Thinking, GPT-5.2 Instant, and GPT-5.2 Thinking) using 3 prompting strategies: a default prompt solely asking the LLMs to classify the case vignettes, a recognition-primed prompt tasking the models to reason according to recognition-primed decision-making, and a data-frame prompt tasking the models to apply the data-frame theory. The task was taken from a standardized and validated evaluation framework and instructed the LLMs to advise on the appropriate care-seeking action for 45 real patient case vignettes across 3 urgency levels (emergency, nonemergency, and self-care). Each model-vignette-prompt combination was tested 10 times to assess and account for output variability. Accuracy was analyzed using mixed effects logistic regression. Additionally, we evaluated accuracy for each urgency level and examined output variability.

Results: Both NDM-inspired prompts increased overall model accuracy (recognition-primed: 67.6%; data-frame: 66.7%) compared to the default prompt (63.3%). The greatest improvements were observed for self-care recommendations, where accuracy increased from 13.4% (default prompt) to 29.8% (recognition-primed prompt) and 24.6% (data-frame prompt). Performance on 2 emergency and 30 nonemergency cases remained high across all prompts. Notably, NDM-inspired prompts made nonreasoning models start giving self-care advice, even though they rarely or never provided self-care advice with the default prompt. Output variability was similar across the 3 prompts.

Conclusions: Using LLMs with prompts inspired by NDM, which are designed to reflect real-world human reasoning, improves the accuracy of LLMs in care-seeking tasks, particularly for self-care advice, without reducing performance in the included emergency or nonemergency cases. These findings indicate that NDM-inspired prompts can offer an advantage when LLMs are used for real-world decisions involving ambiguity and uncertainty. The impact of output that reflects real-world human reasoning on users' decision-making must be evaluated in future studies.

(JMIR Biomed Eng 2026;11:e88053) doi:10.2196/88053

KEYWORDS

prompting; human-technology interaction; human factors; artificial intelligence; decision-making; naturalistic decision-making; naturalistic decision support; cognitive science; care-seeking; self-triage; bounded rationality

Introduction

Since their public release in 2022, large language models (LLMs), such as ChatGPT, have become widely used across domains for a range of tasks [1-8]. Although these models now reach high levels of accuracy on several benchmark tests, both researchers and users are increasingly interested in techniques to further improve model performance through specific input instructions—a process known as "prompting" [9]. Common approaches include assigning the model a specific role, providing relevant context or examples, or specifying a clear output format [10-12]. Three basic prompting strategies are often described in the literature: zero-shot, one-shot, and few-shot prompting. Zero-shot prompting refers to providing only the task instructions without any example outputs. One-shot prompting includes a single example of the expected output, and few-shot prompting provides multiple examples of the expected output [12]. In recent work, prompting strategies focus more on guiding the model through a reasoning process rather than simply providing information. A recent systematic review identified 58 prompting techniques, which were grouped into 6 categories [13]. In addition to zero-shot and few-shot approaches, 4 new categories were described: *ensembling prompts* use multiple prompts and aggregate the resulting outputs [13-16]. *Self-criticism prompts* instruct the model to evaluate and critique its own answers before responding [13,17,18]. *Decomposition prompts* instruct LLMs to break down tasks into smaller steps, which are then solved sequentially [13,19,20]. Finally, *thought generation* or *"chain-of-thought"* prompts ask the model to explicitly explain its reasoning as it works through a problem [13,21,22]. Notably, chain-of-thought and reasoning prompts have now been directly integrated into newer models [23]. For example, OpenAI's o-series models (including o1, o3, and o4) are designed to generate a reasoning response before generating a user response [23]. This approach has been shown to improve accuracy across several benchmarks [21,24,25]. Starting with GPT-5, OpenAI also introduced a new, automatically included reasoning engine that consists of several internal expert models to which a user's request is routed [26,27]. The model also automatically determines the reasoning effort needed to answer the user's request [26,27].

Although the chain-of-thought prompting strategy is inspired by human reasoning, particularly deductive decision-making, there is an ongoing debate about whether LLMs really replicate human reasoning or simply generate plausible-sounding explanations [28]. Shojaee et al [29] recently tested chain-of-thought reasoning models on increasingly complex puzzles and found that LLMs do not engage in consistent reasoning across similar problems. In response, Lawsen [30] argues that these findings can be attributed to experimental artifacts and that LLMs are indeed capable of reasoning consistently and accurately when experimental setups are properly designed. Regardless of whether LLMs truly mimic human deductive reasoning when prompted with reasoning techniques, using human decision-making as a source of inspiration for developing prompting strategies is a promising direction. This is especially true in situations with high uncertainty, where deductive reasoning quickly reaches its limits, but humans are nonetheless able to make fairly good decisions [31].

Insights from the fields of applied psychology and human factors and ergonomics (HF/E) suggest that there is a gap between how humans reason in real-world situations and the assumed standard reasoning approaches related to deduction and induction, which are typically used to instruct and evaluate LLMs [28]. One explanation for this difference may be that humans often make decisions under uncertainty, with incomplete or ambiguous information, and decision tasks and goals are often ill-structured [32-34]. In contrast, most psychology experiments, as well as the current benchmarks for LLMs, rely on well-structured, multiple-choice tasks where all necessary information is explicitly provided, and the LLMs are merely asked to choose the correct answer out of multiple options, which can be readily evaluated against a clear-cut gold standard solution [35-39]. This test format is also used in most educational assessments, and LLMs perform well on this format—for example, passing professional and board certification exams in medicine, psychotherapy, and law. As a result, LLMs are widely promoted as accurate decision-support tools for well-structured tasks, and many users use them for this purpose [37,40-45]. However, when existing models are evaluated on real-world datasets, which more accurately reflect the complexity and ambiguity in real decision-making, their performance seems to be considerably worse [2,35,46].

The distinction between decision-making in idealized situations and in complex, ill-defined real-world settings has long been recognized in behavioral economics and psychology. In 1955, Herbert Simon introduced the concept of bounded rationality to study how human decision-making takes place under limited cognitive and environmental resources rather than under the conditions of perfect knowledge and unlimited resources, the normative ideal of full rationality, which is currently assumed in many LLM benchmarks [47,48]. Building on the idea of bounded rationality, the field of naturalistic decision-making (NDM) developed to study how experts make good decisions in real-world contexts [34,49]. Research in NDM shows that, rather than exhaustively comparing all possible options, both experts and novices typically rely on a limited set of information to recognize the most promising option or action [34,50,51]. This strategy is not perfectly accurate in all situations, but it often results in highly accurate decisions within short timeframes [50,52,53]. Specifically, findings from the NDM field suggest that in situations with high information validity, experts often perform on par with complex algorithms, even when using less information and simpler strategies [31,54,55]. This can be explained by the fact that experts rarely follow a strictly deductive or inductive process. Instead, they quickly recognize which information is relevant and engage in *abductive reasoning*; that is, they generate an initial hypothesis based on the observed piece of relevant information and then seek out information to test this hypothesis and update it as new contradicting information becomes available [28,56].

To describe human decision-making in real-world scenarios, 2 models feature most prominently in the NDM literature: recognition-primed decision-making (RPD) [56] and the data-frame theory [57]. The RPD model, which is used to make quick decisions in familiar situations, consists of 2 core processes: a pattern-matching loop and a mental simulation loop [56]. In the pattern matching loop, decision-makers assess whether a situation is familiar (eg, they recognize whether they have experienced a similar situation before). If the situation is recognized as familiar, they directly implement an action. If it is unfamiliar, they either reassess the situation or seek additional information until they achieve some sense of familiarity and can proceed to the mental simulation loop. In the subsequent mental simulation loop, decision-makers simulate implementing their chosen course of action. If they conclude that this action will most likely work, they implement it; if not, they modify the plan and reassess, or they consider new actions entirely [56].

The data-frame theory focuses more on sense-making and understanding new, unknown situations rather than on making decisions [57]. According to this model, humans use frames (basic ideas, hypotheses, or mental models about what is happening in a given context) and data (information in the environment). These 2 concepts interact: frames determine what information is noticed, sought, and how it is interpreted. At the same time, new data can lead a person to elaborate, revise, or even abandon their current frame. For example, in medical diagnosis, a physician may form an initial hypothesis (frame) based on presenting symptoms, then gather additional data to either confirm or reconsider that frame based on new information [57].

Although there is strong evidence supporting the occurrence, efficiency, and effectiveness of NDM models, such as the RPD model or the data-frame theory in real-world decision-making, these approaches have not yet been applied to instruct LLMs [58,59]. Existing reasoning prompts and models are inspired by an ideal form of human decision-making and deductive reasoning, and they seem to perform well on well-structured problems with known risks and gold-standard solutions, and less so in situations involving real-world ambiguity and uncertainty. Although for the latter situation, NDM-based strategies may prove more effective, they have not yet been applied to improve and evaluate LLM performance in ill-defined, real-world tasks. In this study, we aimed to test whether prompts based on NDM principles can improve LLM performance on a real-world, ill-defined task.

Building on our previous work, we used a standardized and validated evaluation framework that used a care-seeking or “self-triage” decision scenario involving real patient cases to evaluate NDM-based prompts [60]. Self-triage refers to the decision-making process by which people determine whether medical care is needed and, if so, where and how urgently to seek it (eg, self-care, primary care, or emergency care) [6,61,62]. This is a common decision task in everyday life, with 80% to 90% of the population reporting at least 1 symptom within a given month [63,64], and laypeople are increasingly consulting digital tools, such as LLMs, for advice [65,66]. Previous research shows that human performance in self-triage decisions is moderate and that LLMs perform only slightly better on average, although they almost always recommend professional care rather than self-care [6,67-69]. Self-triage is thus a suitable use case for the present study because it is typically ill-structured: information about symptoms may be incomplete and ambiguous, and decision-makers must decide under uncertainty. Therefore, self-triage is a representative example of the real-world decisions studied in NDM research.

We hypothesized that prompts inspired by the RPD model and the data-frame theory will significantly increase accuracy on these tasks in selecting the best course of action across both nonreasoning and reasoning models, compared to a standard zero-shot prompt.

Methods

Study Design

This evaluation study was designed as a prospective, longitudinal, observational LLM validation study. The intervention was the specific prompting strategy: a regular prompt, a recognition-primed prompt, and a data-frame prompt. We used these 3 prompting strategies to assess 45 vignettes across 10 models, each tested 10 times. The primary outcome was the accuracy of the models under each prompting condition, and the secondary outcome was the output variability of the tested models. No participants were involved in this study.

Ethical Considerations

This study did not involve any prospective recruitment, interaction, or intervention with human participants. The LLM evaluation used an existing dataset of symptom descriptions originally collected on an online platform. Ethical approval for this collection, use, and deidentification of the cases was obtained from the ethics committee of the Department of Psychology and Ergonomics at Technische Universität Berlin (AWB_KOP_2_230711). For the present study, we accessed only the deidentified version of these cases [70]. Pseudonymized identifiers (eg, user names) were completely removed, and potential quasi-identifiers in free text (eg, city or institution names) were deleted. The dataset was stored on an access-restricted institutional computer, and the data were used solely for model evaluation. No attempts were made to reidentify individuals. Accordingly, no additional ethical approval was required for this secondary analysis. The reporting of this manuscript follows the TRIPOD (Transparent Reporting of a Multivariable Model for Individual Prognosis or Diagnosis)–LLM guideline [71].

Tested Models

Because ChatGPT remains the most widely used LLM family [72], we focused our evaluation on LLMs currently available within ChatGPT. For the initial data collection, this included GPT-4o, GPT-4.1, GPT-4.1 mini, o3, o4 mini, and o4 mini high. For a second round of data collection, this included GPT-5.1 and GPT-5.2, including both the Instant and Thinking versions. All models are based on the Transformer architecture; however, o3, o4 mini, o4 mini high, GPT-5.1 Thinking, and GPT-5.2 Thinking include a reasoning process prior to generating output for users [23]. All models were tested using the default parameters to approximate consumer-facing ChatGPT model behavior. Thus, we used each model’s default temperature (1), a top-p of 1, did not specify a maximum output length (max_tokens unset), and did not specify a random seed. Because outputs are stochastic without a seed, we repeated each vignette-model-prompt condition 10 times and reported variability to approximate consumer-facing behavior. For GPT-5.1 and GPT-5.2, we set the reasoning.effort parameter to none (to disable reasoning and emulate consumer-facing GPT-5.1 and GPT-5.2 Instant), and to medium (to emulate GPT-5.1 and GPT-5.2 Thinking). Additionally, we conducted a sensitivity analysis using the 2 models for which the prompts yielded the largest accuracy gains. We tested a temperature of 0 (maximum determinism/control) and did not include a higher temperature option because the models frequently refused to provide recommendations with a temperature higher than the default temperature.

Task and Evaluation Dataset

Our task consisted of obtaining advice on which care-seeking option is most appropriate for the described symptoms. This type of task is commonly used to evaluate both digital health applications and LLMs [6,61,62,73-75]. We selected it specifically because it reflects a real use case for ChatGPT [68,76-80], involves uncertainty (ie, unknown risks), and often deals with ambiguous or incomplete data or symptoms. For these reasons, we considered this task well suited to test the influence of NDM-inspired prompts on LLM performance in real-world problems.

The dataset was developed in previous studies and followed current guidelines for evaluating care-seeking decision support systems [70,81-83]. From an “ask the doctor” online platform, 45 real patient cases, where medical laypeople described their symptoms and sought advice from professionals, were collected between October 2023 and January 2024 and psychometrically validated [70,81]. Cases were further stratified to reflect the natural base rates of symptom types that are typically entered into online care-seeking advice tools based on the Centers for Disease Control and Prevention’s National Ambulatory Medical Care Survey [70,84]. Because of this stratification, the dataset included only 2 emergency care cases, 30 nonemergency care cases, and 13 self-care cases. To minimize editing effects, only typos were corrected. The original cases cannot be reproduced in this manuscript for copyright reasons, but they are available from the authors upon reasonable request. The cases describe acute symptoms for which laypeople seek decision support on whether and where to seek care. They cover a range of physical symptom presentations across specialties and are written in natural, nontechnical language from the perspective of medical laypeople. For example, 1 case describes a person reporting pain near the chest or lung area during the night when lying down, which improves when upright but worsens with breathing. After a few minutes, the person noticed a weird sensation, and the pain went away. The person then asks whether they should still seek medical care. Another exemplary case describes a person who notices a visible dent in the lower leg or heel area after sitting with one leg crossed for about 20 minutes. The person says this happens often and is usually painless, but this time the legs seem more swollen than usual, so the dent lasts longer.

To determine the most appropriate level of care, a physician panel consisting of 2 licensed physicians independently rated each case and then discussed their assessments until they reached a consensus assignment on 3 possible care-seeking options: (1) emergency care, defined as situations requiring emergency department visits; (2) nonemergency care, where consulting a general practitioner or another medical specialist is appropriate; and (3) self-care, where patients can monitor or manage symptoms themselves. Further details on the dataset are available in our publication on the RepVig framework [70].

Since the prompts included a specific output format, we used text pattern–based matching to extract the recommendation given by the LLM. If no direct match was found, we assessed the cases manually.

Prompts

To evaluate the baseline accuracy of the tested models, we used the prompt shown in Textbox 1. This prompt was developed in previous studies evaluating the care-seeking advice of LLMs [67,69,70,85] and was adapted to specify an output structure (ie, “Answer: Classification”) that enabled us to classify model responses using natural language processing.

The NDM-inspired prompts were based on this base prompt but additionally included the name of the respective model (ie, “Recognition-Primed Decision-Making model” and “Data-Frame-Theory approach”) to prime the LLMs accordingly. Each prompt then provided instructions on how to reason following the principles of the corresponding model. The full prompts are shown in Textbox 2.

Textbox 1. Care-seeking advice prompt used in the control group to determine the base accuracy of the tested large language models.

Please give a triage classification for the following vignette. The criteria for triage classifications are presented below:

- Emergency: Call 911 or go directly to the emergency room;
- Nonemergency: Try to see a doctor, but the symptoms do not require immediate attention at an emergency room. For example, you can see a doctor in the next week.
- Self-care: Let the health issue get better on its own and review the situation again in a few days.

Respond in the following format:Answer: Your answer [either Emergency, Nonemergency, or Self-Care].

Textbox 2. Recognition-primed prompt instructing the large language models to reason according to the recognition-primed decision-making model.

Please give a triage classification for the following vignette. The criteria for triage classifications are presented below:

- Emergency: Call 911 or go directly to the emergency room;
- Nonemergency: Try to see a doctor, but the symptoms do not require immediate attention at an emergency room. For example, you can see a doctor in the next week.
- Self-care: Let the health issue get better on its own and review the situation again in a few days.

Use the recognition-primed decision-making model to make your decision. Does this situation match any typical cases you know? If yes, what is the usual decision for such a case? Simulate implementing this decision for the described situation and test whether it will work. If not, modify it and test again whether it will work. If the situation is not similar to any typical case you know, try to reassess the situation until you think it sounds familiar, simulate the implementation again, and test whether it will work.

Respond in the following format:Analysis: Your analysis || Answer: Your answer [either Emergency, Nonemergency, or Self-Care].

Textbox 3. Data-frame prompt instructing the large language models to reason according to the data-frame theory.

Please give a triage classification for the following vignette. The criteria for triage classifications are presented below:

- Emergency: Call 911 or go directly to the emergency room;
- Nonemergency: Try to see a doctor, but the symptoms do not require immediate attention at an emergency room. For example, you can see a doctor in the next week.
- Self-care: Let the health issue get better on its own and review the situation again in a few days.

Before giving your triage classification, think about the correct classification using the Data-Frame-Theory approach. As you analyze the vignette, actively use the following reasoning processes (as needed, not necessarily in order):

- Construct or recognize a frame: Identify the main interpretation or mental model that organizes the case information.
- Elaborate the frame: Seek out or infer additional relevant details from the vignette.
- Question the frame: Look for inconsistencies, surprising data, or violated expectations.
- Preserve the frame: Consider whether your interpretation still fits, or if any data needs to be reinterpreted.
- Seek a new frame: If appropriate, consider alternative interpretations.
- Reframe: Revise your perspective and reinterpret the data if needed.
- Compare frames: Identify and weigh alternative ways of understanding the case.

Respond in the following format:Reflection process: Your reflection || Answer: Your answer [either Emergency, Nonemergency, or Self-Care].

All prompts were tested for feasibility in a pretest, during which the authors tested the prompts and API calls using random cases and manually assessed the output for adherence to the instructions and correct formatting.

Procedure

We used a custom-built Python script to access the OpenAI API on May 23, 2025, and a second time for newer models on February 23, 2026. For each model, the prompts (Textboxes 1-3) were entered as system prompts, and the case vignettes as user prompts. The context window was cleared before every call. Because of the high output variability observed in LLMs [67,86,87], we tested each model on each case 10 times as a quality-management measure to account for the fact that different users may receive different advice for the same input. Model outputs were then classified automatically in R into 3 categories (emergency, nonemergency, self-care). For cases in which the category could not be determined through keyword or pattern matching (n=61), manual coding was performed by reading through the answer and assigning a classification manually.

Outcome Measures

The primary outcome was classification accuracy, defined as whether the model's triage recommendation matched the physician-panel gold standard reference for each vignette (ie, correct or incorrect). This metric was chosen because it most closely measures the potential behavioral and safety impact the prompts can have on users. Secondary outcomes included the accuracy by each triage level (emergency, nonemergency, or self-care), calculated as the proportion of correct recommendations within each stratum. Additionally, because the vignette set included only 2 emergency cases, we dichotomized the triage levels into 2 groups: requiring medical care versus self-care. Next, we assessed output variability using Fleiss' Kappa for each model-vignette-prompt combination, and by assessing the consistency of model recommendations, that is, the proportion of the 10 trials corresponding to the most frequently given recommendation. Lastly, we assessed technical accuracy by coding whether the correct recommendation was given at least once among all 10 trials.

Data Analysis

All analyses were conducted in R using the packages symptomcheckR, tidyverse, psych, and lme4 [88-91]. To assess the accuracy of each prompt, we calculated the mean proportion of correctly solved cases and quantified precision using 95% CIs. To test our hypothesis that the NDM-inspired prompts increase LLM performance, we used mixed effects binomial logistic regression (with prompt type as a fixed effect and random intercepts for model, vignette, and model-by-vignette combination to account for repeated observations and clustering in our data—that is, for having each model assess each vignette 10 times) with 2-sided tests. Additionally, we conducted subgroup analyses to assess accuracy in dichotomized decisions (ie, professional care vs self-care), accuracy by model, and accuracy by each care-seeking level. In sensitivity analyses, we further tested whether the reported results remained stable with a low-temperature setting. We chose the 2 models with the highest and lowest prompt-dependent accuracy improvement (ie, GPT-4.1 mini and GPT-5.2 Instant).

To quantify output variability for each prompt, vignette, and model combination, we calculated Fleiss κ and recorded the frequency with which the most common recommendation was given across the 10 trials for each vignette and model. As an estimate of “technical accuracy” (ie, whether the model was technically capable of generating the correct advice), we noted whether the correct recommendation was given at least once in 10 trials [67,92].

Results

Assessments

We used 45 vignettes to test 10 models, with each model run 10 times per vignette using 3 prompting strategies (default, recognition-primed prompting, and data-frame prompting). This resulted in a total of 13,500 individual assessments.

Overall Accuracy of Each Prompt

The average accuracy across all models, vignettes, and trials was 63.3% (95% CI 61.9% - 64.7%) for the default prompt, 67.6% (95% CI 66.3% - 69%) for the recognition-primed prompt, and 66.7% (95% CI 65.3% - 68%) for the data-frame prompt. Both the recognition-primed prompt (OR 2.26, z=8.69; P<.001) and the data-frame prompt (OR 2.05, z=7.23; P<.001) significantly increased accuracy compared to the default prompt. Improvements were greater for reasoning models than for nonreasoning models (OR 2.15, z=4.05, P<.001 for the recognition-primed prompt and OR 1.70, z=2.65, P=.008 for the data-frame prompt). The largest increase in accuracy compared to the default prompt was observed for GPT-4.1 mini with the data-frame prompt, with an improvement of 13 percentage points (95% CI 9.7 - 16.1), as shown in Table 1. In a binary choice, that is, care versus self-care, results remained similar, as shown in Table S1 in Multimedia Appendix 1. The same holds true when tested with a low-temperature setting, as shown in Table S2 in Multimedia Appendix 1.

Table . The accuracy of all tested models for each prompt.

Model	Default prompt, mean (95% CI)	Recognition-primed prompt, mean (95% CI)	Data-frame prompt, mean (95% CI)	Model type
Overall (%)	63.3 (61.9 - 64.7)	67.6 (66.3 - 69)	66.7 (65.3 - 68)	—[a]
GPT-4o (%)	65.3 (60.8 - 69.6)	70.7 (66.3 - 74.7)	66.2 (61.7 - 70.4)	Nonreasoning
GPT-4.1 (%)	64.4 (59.9 - 68.7)	72.7 (68.4 - 76.6)	72.4 (68.1 - 76.4)	Nonreasoning
GPT-4.1 mini (%)	49.8 (45.2 - 54.4)	60.7 (56.1 - 65.1)	62.4 (57.9 - 66.8)	Nonreasoning
o3 (%)	70.7 (66.3 - 74.7)	75.1 (70.9 - 78.9)	76.2 (72.1 - 79.9)	Reasoning
o4 mini (%)	69.3 (64.9 - 73.4)	70.7 (66.3 - 74.7)	72.9 (68.6 - 76.8)	Reasoning
o4 mini high (%)	68.9 (64.5 - 73)	71.3 (67 - 75.3)	70.7 (66.3 - 74.7)	Reasoning
GPT-5.1 Instant (%)	64 (59.4 - 68.4)	71.1 (66.7 - 75.3)	66.2 (61.6 - 70.6)	Nonreasoning
GPT-5.1 Thinking (%)	70.7 (66.2 - 74.8)	74.7 (70.4 - 78.6)	72.4 (68.1 - 76.5)	Reasoning
GPT-5.2 Instant (%)	57.8 (53.1 - 62.4)	56.4 (51.7 - 61.1)	55.3 (50.6 - 60)	Nonreasoning
GPT-5.2 Thinking (%)	52 (47.3 - 56.7)	53.1 (48.4 - 57.8)	51.8 (47.1 - 56.5)	Reasoning

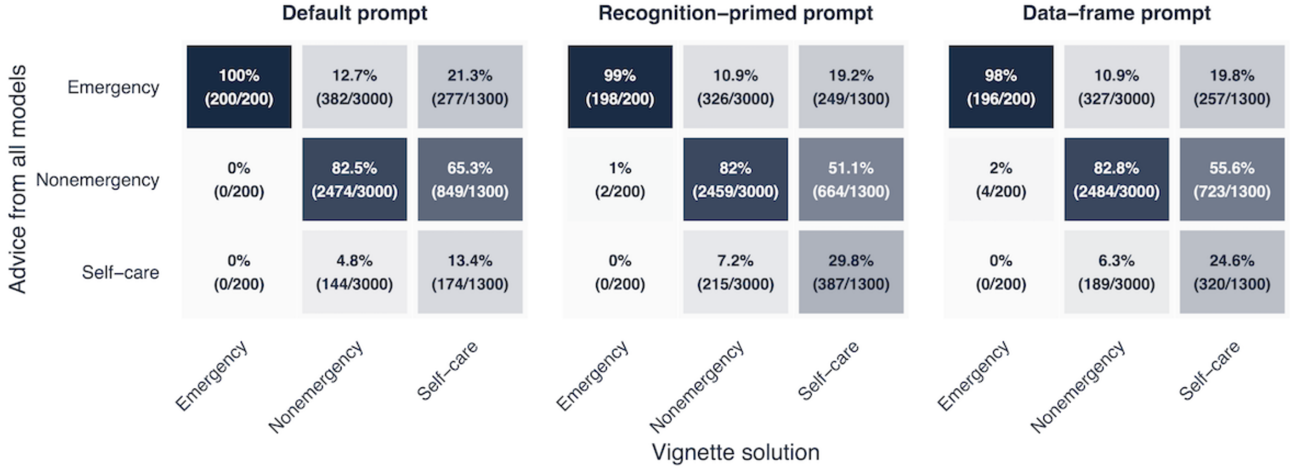
[a]Not available.

The recognition-primed prompt increased accuracy in 18 of 45 cases (40%) and decreased accuracy in 13 cases (29%) on average across all models. Its median increase in accuracy was 18% (IQR 5% - 24%), whereas the median decrease was 5% (IQR 2% - 7%). The data-frame prompt increased accuracy in 17 of 45 (38%) cases and reduced it in 12 (27%) cases. Its median increase in accuracy was 14% (IQR 4% - 19%), and the median decrease was 5% (IQR 3% - 8%). Most decreases in accuracy affected nonemergency cases (12/13, 92% for the recognition-primed prompt and 11/12, 92% for the data-frame prompt, see the Triage-Level Accuracy of Each Prompt section for the direction). Increases were observed in both nonemergency and self-care cases (8/18, 44%, and 10/18, 56%, respectively, for the recognition-primed prompt; 8/17, 47%, and 9/17, 53%, for the data-frame prompt, see the next section for the direction) (Table S3 in Multimedia Appendix 1).

Triage-Level Accuracy of Each Prompt

Across all 3 prompts, the models tended to recommend higher-than-necessary urgency (accounting for 88% of all errors, 95% CI 87% - 88.9%) rather than lower-than-necessary urgency (12% of all errors, 95% CI 11.1% - 13%). With the default prompt, both emergency cases were correctly identified (100%, 95% CI 98.2% - 100%). Using both the data-frame prompt and the recognition-primed prompt, both emergency cases were also mostly identified correctly (99%, 95% CI 96.4% - 99.9% and 98%, 95% CI 95% - 99.5%, respectively), although some trials resulted in incorrect nonemergency advice (1%, 95% CI 0.1% - 3.6% and 2%, 95% CI 0.5% - 5%, respectively). Accuracy for nonemergency cases was similar across all prompts: 82.5% (95% CI 81.1% - 83.8%) for the default prompt, 82% (95% CI 80.5% - 83.3%) for the recognition-primed prompt, and 82.8% (95% CI 81.4% - 84.1%) for the data-frame prompt. The largest difference was observed for self-care cases: With the default prompt, the models correctly identified only 13.4% (95% CI 11.6% - 15.4%), compared to 29.8% (95% CI 27.3% - 32.3%) with the recognition-primed prompt and 24.6% (95% CI 22.3% - 27.1%) with the data-frame prompt (Figure 1). The results remained similar in a binary choice task and also when tested with a low temperature setting (Tables S1 and S4 in Multimedia Appendix 1).

Figure 1. Confusion matrix showing the classification of each prompt across all models compared to the correct vignette solution. Emergency estimates may be unreliable because only 2 cases were included.



Notably, nonreasoning models that never or rarely provided self-care advice with the default prompt began providing self-care advice with relatively high accuracy when using the NDM-inspired prompts (eg, 0%, 95% CI 0% - 2.9% for the default prompt in GPT-4.1, compared to 43.8%, 95% CI 35.6% - 52.4% for the recognition-primed prompt and 39.2%, 95% CI 31.3% - 47.8% for the data-frame prompt). For reasoning models, which already gave self-care advice with the default prompt, accuracy further improved with the NDM-inspired prompts (eg, 46.9%, 95% CI 38.6% - 55.5% with the default prompt in o4 mini; 63.8%, 95% CI 55.3% - 71.6% with the recognition-primed prompt; and 56.2%, 95% CI 47.6% - 64.4% with the data-frame prompt; Table 2).

Table . Accuracy of each model and prompt in generating care-seeking advice by correct vignette solution.

Model and vignette type	Default prompt, mean (95% CI)	Recognition-primed prompt, mean (95% CI)	Data-frame prompt, mean (95% CI)
GPT-4o (%)			
Emergency (n=2)	100 (83.9 - 100)	100 (83.9 - 100)	100 (83.9 - 100)
Nonemergency	90.3 (86.5 - 93.2)	92.3 (88.8 - 94.8)	87.7 (83.5 - 90.9)
Self-care	2.3 (0.8 - 6.6)	16.2 (10.8 - 23.4)	11.5 (7.1 - 18.2)
GPT-4.1 (%)			
Emergency (n=2)	100 (83.9 - 100)	100 (83.9 - 100)	80 (58.4 - 91.9)
Nonemergency	90 (86.1 - 92.9)	83.3 (78.7 - 87.1)	86.3 (82 - 89.8)
Self-care	0 (0 - 2.9)	43.8 (35.6 - 52.4)	39.2 (31.3 - 47.8)
GPT-4.1 mini (%)			
Emergency (n=2)	100 (83.9 - 100)	100 (83.9 - 100)	100 (83.9 - 100)
Nonemergency	68 (62.5 - 73)	81.3 (76.5 - 85.3)	87 (82.7 - 90.3)
Self-Care	0 (0 - 2.9)	6.9 (3.7 - 12.6)	0 (0 - 2.9)
o3 (%)			
Emergency (n=2)	100 (83.9 - 100)	100 (83.9 - 100)	100 (83.9 - 100)
Nonemergency	93.7 (90.3 - 95.9)	90.7 (86.8 - 93.5)	92.7 (89.1 - 95.1)
Self-care	13.1 (8.3 - 19.9)	35.4 (27.7 - 43.9)	34.6 (27 - 43.1)
o4 mini (%)			
Emergency (n=2)	100 (83.9 - 100)	100 (83.9 - 100)	100 (83.9 - 100)
Nonemergency	77 (71.9 - 81.4)	71.7 (66.3 - 76.5)	78.3 (73.3 - 82.6)
Self-care	46.9 (38.6 - 55.5)	63.8 (55.3 - 71.6)	56.2 (47.6 - 64.4)
o4 mini high (%)			
Emergency (n=2)	100 (83.9 - 100)	100 (83.9 - 100)	100 (83.9 - 100)
Nonemergency	75.7 (70.5 - 80.2)	74.7 (69.5 - 79.3)	74 (68.8 - 78.6)
Self-care	48.5 (40 - 57)	59.2 (50.6 - 67.3)	58.5 (49.9 - 66.6)
GPT-5.1 Instant (%)			
Emergency (n=2)	100 (83.9 - 100)	95 (76.4 - 99.1)	100 (83.9 - 100)
Nonemergency	87 (82.7–90.3)	86 (81.6 - 89.5)	88 (83.8 - 91.2)
Self-care	5.4 (2.6 - 10.7)	33.1 (25.6 - 41.5)	10.8 (6.5 - 17.3)
GPT-5.1 Thinking (%)			
Emergency (n=2)	100 (83.9 - 100)	95 (76.4 - 99.1)	100% (83.9% - 100%)
Nonemergency	92 (88.4 - 94.6)	90 (86.1 - 92.9)	87 (82.7 - 90.3)
Self-care	16.9 (11.4 - 24.3)	36.2 (28.4 - 44.7)	34.6 (27 - 43.1)
GPT-5.2 Instant (%)			
Emergency (n=2)	100 (83.9 - 100)	100 (83.9 - 100)	100 (83.9 - 100)
Nonemergency	80 (75.1 - 84.1)	76.7 (71.6 - 81.1)	76.3 (71.2 - 80.8)
Self-care	0 (0 - 2.9)	3.1 (1.2 - 7.6)	0 (0 - 2.9)
GPT-5.2 Thinking (%)			
Emergency (n=2)	100 (83.9 - 100)	100 (83.9 - 100)	100 (83.9 - 100)
Nonemergency	71 (65.6 - 75.8)	73 (67.7 - 77.7)	70.7 (65.3 - 75.5)
Self-care	0.8 (0.1 - 4.2)	0 (0 - 2.9)	0.8 (0.1 - 4.2)

Output Variability of Each Prompting Technique

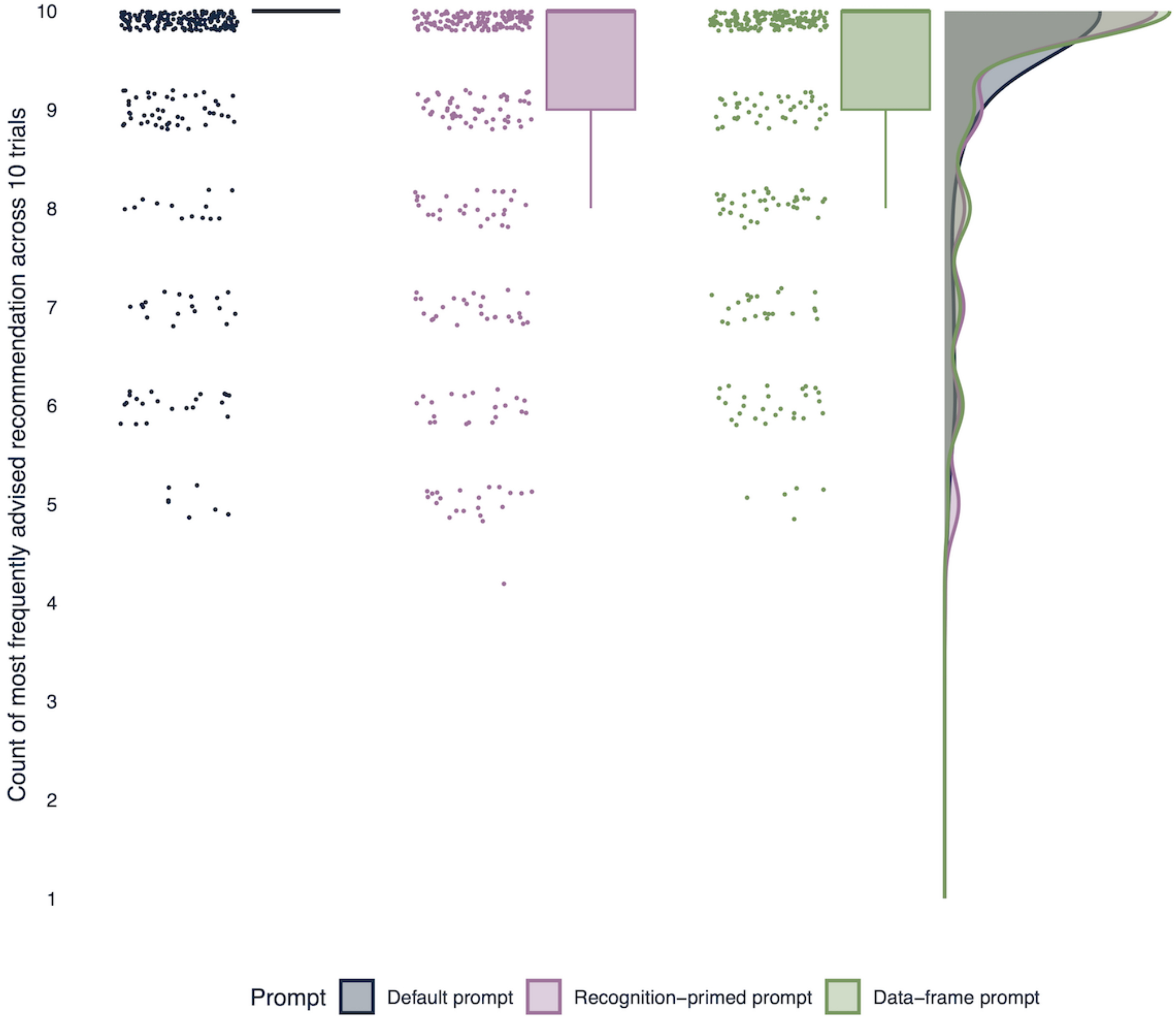
Intertrial reliability—that is, the frequency with which a vignette received the same advice from the same model across multiple trials—was comparable across all prompts, with median Fleiss κ values of 0.766 (IQR 0.706 - 0.884) for the default prompt, 0.717 (IQR 0.681 - 0.743) for the recognition-primed prompt, and 0.751 (IQR 0.725 - 0.773) for the data-frame prompt (Table 3). The results remained similar when tested with a low temperature setting (Table S5 in Multimedia Appendix 1).

Table . Intertrial reliability (Fleiss κ) of each model for each prompt.

Model	Default prompt	Recognition-primed prompt	Data-frame prompt
Overall, median (IQR)	0.766 (0.706 - 0.884)	0.717 (0.681 - 0.743)	0.751 (0.725 - 0.773)
GPT-4o	0.708	0.710	0.718
GPT-4.1	0.963	0.777	0.658
GPT-4.1 mini	0.824	0.634	0.756
o3	0.706	0.675	0.807
o4 mini	0.707	0.699	0.766
o4 mini high	0.701	0.743	0.744
GPT-5.1 Instant	0.895	0.644	0.620
GPT-5.1 Thinking	0.689	0.723	0.745
GPT-5.2 Instant	0.926	0.742	0.775
GPT-5.2 Thinking	0.851	0.821	0.839

When considering the most frequently given recommendation for each vignette by each model, all prompts yielded relatively consistent advice across multiple trials (mean 76.9%, 95% CI 72.7% - 80.7% for the default prompt; mean 66.9%, 95% CI 62.3% - 71.2% for the recognition-primed prompt; mean 71.1%, 95% CI 66.7% - 75.3% for the data-frame prompt), as shown in Figure 2. There was no statistically significant difference between the prompts in how often a specific option was recommended across trials (z=1.50, P=.13 for the recognition-primed prompt; z=0.72, P=.47 for the data-frame prompt).

Figure 2. Number of times the most frequently advised recommendation was given among the 3 prompts.



However, the tested models were more likely to provide the correct solution at least once across multiple trials when using the recognition-primed prompt (mean 82.2%, 95% CI 78.4% - 85.6%) and the data-frame prompt (mean 78.4%, 95% CI 74.4% - 82.2%) compared to the default prompt (mean 73.1%, 95% CI 68.8% - 77.2%) (Table 4). The results remained similar when tested with a low-temperature setting (Table S6 in Multimedia Appendix 1).

Table . Percentage of cases that were solved correctly at least once among 10 trials.

Model	Default prompt, mean (95% CI)	Recognition-primed prompt, mean (95% CI)	Data-frame prompt, mean (95% CI)
Overall (%)	73.1 (68.8 - 77.2)	82.2 (78.4 - 85.6)	78.4 (74.4 - 82.2)
GPT-4o (%)	77.8 (62.9 - 88.8)	84.4 (70.5 - 93.5)	80 (65.4 - 90.4)
GPT-4.1 (%)	66.7 (51 - 80)	84.4 (70.5 - 93.5)	84.4 (70.5 - 93.5)
GPT-4.1 mini (%)	55.6 (40 - 70.4)	75.6 (60.5 - 87.1)	66.7 (51 - 80)
o3 (%)	82.2 (67.9 - 92)	91.1 (78.8 - 97.5)	84.4 (70.5 - 93.5)
o4 mini (%)	86.7 (73.2 - 94.9)	88.9 (75.9 - 96.3)	86.7 (73.2 - 94.9)
o4 mini high (%)	91.1 (78.8 - 97.5)	88.9 (75.9 - 96.3)	86.7 (73.2 - 94.9)
GPT-5.1 Instant (%)	68.9 (53.4 - 81.8)	86.7 (73.2 - 94.9)	82.2 (67.9 - 92)
GPT-5.1 Thinking (%)	82.2 (67.9 - 92)	88.9 (75.9 - 96.3)	86.7 (73.2 - 94.9)
GPT-5.2 Instant (%)	60 (44.3 - 74.3)	73.3 (58.1 - 85.4)	64.4 (48.8 - 78.1)
GPT-5.2 Thinking (%)	60 (44.3 - 74.3)	60 (44.3 - 74.3)	62.2 (46.5 - 76.2)

Discussion

Principal Results

Our study investigated whether prompting strategies inspired by NDM—a field that analyzes how humans make real-world decisions under uncertainty—can improve LLM performance in ill-defined tasks such as care-seeking decisions. Our results show that both the recognition-primed and the data-frame prompts increased the accuracy of care-seeking advice across all tested models except GPT-5.2. Although this effect may partly reflect the additional reasoning process before producing an answer among nonreasoning models, we observed improvements not only in nonreasoning models but also in reasoning models that already include a reasoning process. This observation suggests that our results cannot simply be attributed to a general reasoning process. Notably, most nonreasoning models with NDM-inspired prompts outperformed traditional reasoning models using the default prompt, and reasoning models also showed significant improvements with the NDM-inspired prompts.

The greatest improvements due to the NDM-inspired prompts were seen in self-care cases, which were more often correctly identified. Nonreasoning models rarely or never provided self-care advice with the default prompt, a finding consistent with previous studies [67,69,70]. When prompted with the NDM-inspired prompts, these models began giving self-care advice and even reached a relatively high level of accuracy, up to 44%. In contrast, accuracy for the 2 included emergency cases and the included nonemergency cases showed little change, likely due to a ceiling effect, as the tested models were already highly accurate on these cases with the default prompt. Prior research suggests that self-care advice is typically given by LLMs only when a reasoning process is included and that the tendency toward risk-averse recommendations may stem from built-in safety measures [67,69]. The recognition-primed prompt explicitly instructs the model to recall similar situations (pattern matching according to the RPD model) and to forecast possible outcomes (mental simulation), which may help the model reconsider overly cautious recommendations before giving advice. Similarly, the data-frame prompt encourages the model to re-examine each initial recommendation and—if new data do not fit the initial frame—explore alternative frames, which may help identify when self-care is sufficient rather than defaulting to medical referral.

OpenAI's most recent GPT-5 model family includes an updated reasoning process [26,27]. The benefits of using NDM-inspired prompts were replicated for GPT-5.1 (for both the Instant version without a reasoning process and the Thinking version with a reasoning process), but not for GPT-5.2: self-care accuracy dropped to 0% in both GPT-5.2 Instant and GPT-5.2 Thinking and remained unchanged with NDM-inspired prompts. These results may suggest a version-level shift toward recommending professional care that prompting does not alter. This observation is unlikely to be attributable to changes in the reasoning mechanism alone, because GPT-5.1—despite also using the updated reasoning process—did not show the same decrease in self-care accuracy.

Implications

The present findings have implications for prompt engineering, artificial intelligence (AI) research, and end users. First, for prompt engineering, we suggest that, rather than relying solely on prompts built on computer science (eg, ensemble methods and decomposing), strategies derived from cognitive science, applied psychology, and HF/E—especially those based on models of human decision-making under uncertainty—may be more effective or, at least, serious competitors, particularly in domains with high ambiguity and uncertainty, such as triage or diagnostic decisions. In these ill-defined situations, we showed that a "reasoning blueprint" based on human cognition can outperform methods that simply instruct the models to reason. We acknowledge, however, that, based on our results, the benefits of NDM-inspired prompts are thus far limited to uncertain tasks. It remains to be seen how they perform on more well-defined tasks, such as text formatting or summarization.

Limitations

Although our results show a positive impact of combining NDM with prompt engineering, there are several limitations. First, we conducted a single benchmarking test within one domain, that is, care-seeking advice. Although this is a typical real-world decision task with high uncertainty and a common use case for LLMs [68,76-80], it remains unclear whether our findings generalize to other tasks or domains with varying levels of ambiguity and/or uncertainty. Second, the low sample size for emergency cases leads to unstable accuracy estimates, and no safety conclusions should be derived from the data presented here. Third, we limited our evaluation to LLMs that are currently integrated into ChatGPT. We made this decision to assess practical impact for users; however, it is unclear whether these results would hold true for the broader range of LLMs available or in development. In particular, future research should test whether similar results can be achieved with smaller models and limited context windows, given that the reasoning process increases token requirements.

The NDM-inspired prompts themselves present another limitation. These prompts are computationally more expensive to run than standard user inputs because they add a reasoning output before giving advice. Although this may not affect individual users, it could increase operational costs for developers integrating such prompts, especially compared to nonreasoning models. We recommend that any potential performance gains from NDM-inspired prompting be carefully weighed against increased costs on a case-by-case basis.

Next, we did not include participants who interacted with the LLMs directly. Instead, we used a highly controlled setup in which each model was prompted repeatedly using standardized prompts. In real-world use, however, users' prompts vary substantially in both content and quality [4,70,93]. Accordingly, the present study was designed to test whether NDM-inspired prompts can improve model accuracy under controlled conditions; we cannot infer that these prompts would translate into improved user decisions or higher-quality outputs in everyday use. This work should, therefore, be interpreted as a technical evaluation of model behavior under controlled inputs rather than as a clinical validation study. Depending on the intended use, LLMs may be regulated as Software as a Medical Device and may, therefore, require additional evidence that is outside the scope of the present study. Recent work on differing user inputs and adversarial attacks in chatbots shows that they can produce unsafe outputs depending on the specific prompts, which further demonstrates that more rigorous and use-case-specific safety evaluations are needed before deployment [94,95]. Future studies should therefore conduct user studies to examine whether NDM-inspired prompts also yield better recommendations and decision support for users in real-world settings, and to determine how NDM-inspired prompts may be used to prevent adversarial attacks.

Finally, the prompts tested here were based on only 2 decision-making models. There are other models that could serve as inspiration for prompt development, such as the decision ladder or heuristic decision models [51,96]. Moreover, domain-specific decision-making models may be even better suited for certain use cases. For care-seeking advice, no such model currently exists to explain how humans make these decisions. However, the development of such a model could be helpful to develop even more targeted prompting strategies to further increase LLM performance.

Future Research

This study is among the first to combine NDM and AI-based decision support systems to foster more naturalistic decision support. Our findings provide a foundation for future work by demonstrating that real-world human reasoning strategies can improve the accuracy of LLMs. Building on these results, future work could examine how NDM and AI can be combined to support users. For example, prompting LLMs to use reasoning processes that reflect human decision-making could open a new direction for explainable AI. Unlike traditional explainable AI methods that focus on feature importance, providing explanations based on human-like pattern recognition and mental simulation may increase trust and help users identify potential mistakes in the reasoning process. Prior research has shown that users critically assess, rather than blindly follow, AI advice [4]. Giving users an NDM-inspired reasoning approach may support this evaluation more than providing advice with a post hoc explanation.

NDM-inspired prompts may also improve human-AI collaboration: When humans and AI share a conceptual language (consisting of frames, pattern matching, and mental simulation), it may become easier for users to integrate AI advice into their own reasoning. For example, physicians could review the frames used by the LLM, add new data points, and let the AI simulate whether these fit the frame. Conversely, the AI could make predictions based on its frame, which the physician can cross-check with clinical data. An AI would thus not only give a final recommendation but also provide support in hypothesis generation, data gathering, and hypothesis testing [97,98].

Next, NDM-inspired prompting could also be used for education and training. LLMs could serve as interactive tools for medical students, allowing them to practice decision-making using the RPD model alongside the AI by comparing their mental simulations with those of the model. The AI could then provide feedback on differences in their respective frames.

More broadly, future work should move beyond technical benchmarking toward evaluation designs aligned with Software as a Medical Device expectations by predefining the intended use case and testing performance and safety prospectively in real-world settings. In this context, NDM may be treated as a theoretical basis for uncertainty management, and future studies can test whether NDM-based prompts reduce failures across different user inputs and adversarial attacks.

Conclusions

In this study, we showed that applying models from NDM to prompt LLMs can improve performance in highly uncertain and ambiguous care-seeking tasks. Both NDM-inspired prompts tested here increased overall accuracy across both reasoning and nonreasoning models, with the greatest improvement in self-care recommendations, while maintaining high accuracy in the 2 included emergency cases and all included nonemergency cases. These findings may open up a new strategy for prompt engineering: rather than relying on prompts derived from computer science, prompts that build on NDM models or related models from applied psychology and HF/E, which represent how humans make sense of uncertainty, may be more effective in ill-defined tasks. As LLMs and other AI tools are increasingly adopted in safety-critical and everyday applications, NDM-inspired prompting may offer a strategy for making AI more useful for real-world decision-making.

Acknowledgments

The authors declare the use of generative artificial intelligence (GAI) in the research and writing process. According to the GAIDeT (Generative AI Delegation Taxonomy; 2025), the following tasks were delegated to GAI tools under full human supervision: proofreading and editing. The GAI tool used was ChatGPT, model GPT-4.1. Responsibility for the final manuscript lies entirely with the authors. GAI tools are not listed as authors and do not bear responsibility for the final outcomes.

Funding

The authors declared no financial support was received for this work.

Data Availability

The data can be accessed via Zenodo [99].

Conflicts of Interest

MK is an associate editor for *JMIR Public Health and Surveillance*.

Multimedia Appendix 1
Additional accuracy and sensitivity analyses.
[DOCX File, 30 KB - biomedeng_v11i1e88053_app1.docx]

References

1. Jo E, Song S, Kim JH, et al. Assessing GPT-4's performance in delivering medical advice: comparative analysis with human experts. JMIR Med Educ 2024 Jul 8;10:e51282. [doi: 10.2196/51282] [Medline: 38989848]
2. Rao A, Pang M, Kim J, et al. Assessing the utility of ChatGPT throughout the entire clinical workflow: development and usability study. J Med Internet Res 2023 Aug 22;25:e48659. [doi: 10.2196/48659] [Medline: 37606976]
3. Kisvarday S, Yan A, Yarahuan J, et al. ChatGPT use among pediatric health care providers: cross-sectional survey study. JMIR Form Res 2024 Sep 12;8:e56797. [doi: 10.2196/56797] [Medline: 39265163]
4. Kopka M, Wang SM, Kunz S, Schmid C, Feufel MA. Technology-supported self-triage decision making. NPJ Health Syst 2025 Jan 25;2(1):1-11. [doi: 10.1038/s44401-024-00008-x]
5. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. Nat Med 2023 Aug;29(8):1930-1940. [doi: 10.1038/s41591-023-02448-8] [Medline: 37460753]
6. Kopka M, von Kalckreuth N, Feufel MA. Accuracy of online symptom assessment applications, large language models, and laypeople for self-triage decisions. NPJ Digit Med 2025 Mar 25;8(1):178. [doi: 10.1038/s41746-025-01566-6] [Medline: 40133390]
7. Memarian B, Doleck T. ChatGPT in education: methods, potentials, and limitations. Comput Hum Behav Artif Humans 2023 Aug;1(2):100022. [doi: 10.1016/j.chbah.2023.100022]
8. Ajevski M, Barker K, Gilbert A, Hardie L, Ryan F. ChatGPT and the future of legal education and practice. Law Teach 2023 Jul 3;57(3):352-364. [doi: 10.1080/03069400.2023.2207426]
9. Henrickson L, Meroño-Peñuela A. Prompting meaning: a hermeneutic approach to optimising prompt engineering with ChatGPT. AI Soc 2025 Feb;40(2):903-918. [doi: 10.1007/s00146-023-01752-8]
10. Housley W, Dahl P. Membership categorisation, sociological description and role prompt engineering with ChatGPT. Discourse Commun 2024 Dec;18(6):848-858. [doi: 10.1177/17504813241267068]
11. Meskó B. Prompt engineering as an important emerging skill for medical professionals: tutorial. J Med Internet Res 2023 Oct 4;25(1):e50638. [doi: 10.2196/50638] [Medline: 37792434]
12. Reynolds L, McDonell K. Prompt programming for large language models: beyond the few-shot paradigm. Presented at: Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems (CHI EA '21); May 8-13, 2021. [doi: 10.1145/3411763.3451760]
13. Schulhoff S, Ilie M, Balepur N, et al. The prompt report: a systematic survey of prompt engineering techniques. arXiv. Preprint posted online on Jun 6, 2024. [doi: 10.48550/arXiv.2406.06608]
14. Si C, Shi W, Zhao C, Zettlemoyer L, Boyd-Graber J. Getting more out of mixture of language model reasoning experts. Presented at: Findings of the Association for Computational Linguistics: EMNLP 2023; Dec 6-10, 2023. [doi: 10.18653/v1/2023.findings-emnlp.552]
15. Sorensen T, Robinson J, Rytting C, et al. An information-theoretic approach to prompt engineering without ground truth labels. Presented at: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers); May 22-27, 2022. [doi: 10.18653/v1/2022.acl-long.60]
16. Wang X, Wei J, Schuurmans D, et al. Self-consistency improves chain of thought reasoning in language models. Presented at: Proceedings of the 11th International Conference on Learning Representations (ICLR 2023); May 1-5, 2023 URL: https://openreview.net/pdf?id=1PL1NIMMrw [accessed 2026-03-18]
17. Madaan A, Tandon N, Gupta P. SELF-REFINE: iterative refinement with SELF-feedback. Presented at: Proceedings of the 37th International Conference on Neural Information Processing Systems (NeurIPS 2023); Dec 10-16, 2023 URL: https://openreview.net/pdf?id=S37hOerQLB [accessed 2026-03-18]
18. Weng Y, Zhu M, Xia F, et al. Large language models are better reasoners with self-verification. Presented at: Findings of the Association for Computational Linguistics: EMNLP 2023; Dec 6-10, 2023. [doi: 10.18653/v1/2023.findings-emnlp.167]
19. Patel P, Mishra S, Parmar M, Baral C. Is a question decomposition unit all we need? Presented at: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing; Dec 7-11, 2022. [doi: 10.18653/v1/2022.emnlp-main.302]
20. Wang L, Xu W, Lan Y, et al. Plan-and-solve prompting: improving zero-shot chain-of-thought reasoning by large language models. Presented at: Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers); Jul 9-14, 2023. [doi: 10.18653/v1/2023.acl-long.147]
21. Wei J, Wang X, Schuurmans D. Chain-of-thought prompting elicits reasoning in large language models. Presented at: NIPS'22: Proceedings of the 36th International Conference on Neural Information Processing Systems; Nov 28 to Dec 9, 2022 URL: https://dl.acm.org/doi/10.5555/3600270.3602070 [accessed 2026-03-18]

22. Besta M, Blach N, Kubicek A, et al. Graph of thoughts: solving elaborate problems with large language models. Presented at: Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence and Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence and Fourteenth Symposium on Educational Advances in Artificial Intelligence; Feb 27 to Mar 3, 2024. [doi: 10.1609/aaai.v38i16.29720]
23. Reasoning models. OpenAI Developers. URL: https://developers.openai.com/api/docs/guides/reasoning [accessed 2025-06-27]
24. Gu SS, Iwasawa Y, Kojima T, Matsuo Y, Reid M. Large language models are zero-shot reasoners. Presented at: NIPS'22: Proceedings of the 36th International Conference on Neural Information Processing Systems; Nov 28 to Dec 9, 2022. [doi: 10.52202/068431-1613]
25. Shah HA, Househ M. Chain of thought strategy for smaller LLMs for medical reasoning. Stud Health Technol Inform 2025 May 15;327:783-787. [doi: 10.3233/SHTI250466] [Medline: 40380574]
26. Leon M. GPT-5 and open-weight large language models: advances in reasoning, transparency, and control. Inf Syst 2026 Feb;136:102620. [doi: 10.1016/j.is.2025.102620]
27. Wang S, Hu M, Li Q, Safari M, Yang X. Capabilities of GPT-5 on multimodal medical reasoning. arXiv. Preprint posted online on Aug 11, 2025. [doi: 10.48550/ARXIV.2508.08224]
28. Flach JM, Feufel MA, Reynolds PL, Parker SH, Kellogg KM. Decisionmaking in practice: the dynamics of muddling through. Appl Ergon 2017 Sep;63:133-141. [doi: 10.1016/j.apergo.2017.03.017] [Medline: 28502402]
29. Shojaee P, Mirzadeh I, Alizadeh K, Horton M, Bengio S, Farajtabar M. The illusion of thinking. SI 2025;2(6) [FREE Full text] [doi: 10.70777/si.v2i6.15919]
30. Lawsen A. Comment on the illusion of thinking: understanding the strengths and limitations of reasoning models via the lens of problem complexity. arXiv. Preprint posted online on Jun 10, 2025. [doi: 10.48550/arXiv.2506.09250]
31. Kahneman D, Klein G. Conditions for intuitive expertise: a failure to disagree. Am Psychol 2009 Sep;64(6):515-526. [doi: 10.1037/a0016755] [Medline: 19739881]
32. Samuel AN, Peter MT. Ecological rationality: bounded rationality in an evolutionary light. In: Viale R, editor. Routledge Handbook of Bounded Rationality, 1st edition: Routledge; 2020:280-312.
33. Dhami MK, Hertwig R, Hoffrage U. The role of representative design in an ecological approach to cognition. Psychol Bull 2004 Nov;130(6):959-988. [doi: 10.1037/0033-2909.130.6.959] [Medline: 15535744]
34. Klein G. Naturalistic decision making. Hum Factors 2008 Jun;50(3):456-460. [doi: 10.1518/001872008X288385] [Medline: 18689053]
35. Hager P, Jungmann F, Holland R, et al. Evaluation and mitigation of the limitations of large language models in clinical decision-making. Nat Med 2024 Sep;30(9):2613-2622. [doi: 10.1038/s41591-024-03097-1] [Medline: 38965432]
36. Singhal K, Azizi S, Tu T, et al. Large language models encode clinical knowledge. Nature New Biol 2023 Aug;620(7972):172-180. [doi: 10.1038/s41586-023-06291-2] [Medline: 37438534]
37. Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLoS Digit Health 2023 Feb;2(2):e0000198. [doi: 10.1371/journal.pdig.0000198] [Medline: 36812645]
38. Kanjee Z, Crowe B, Rodman A. Accuracy of a generative artificial intelligence model in a complex diagnostic challenge. JAMA 2023 Jul 3;330(1):78-80. [doi: 10.1001/jama.2023.8288] [Medline: 37318797]
39. Eriksen AV, Möller S, Ryg J. Use of GPT-4 to diagnose complex clinical cases. NEJM AI 2024 Jan;1(1):AIp2300031. [doi: 10.1056/AIp2300031]
40. Jung LB, Gudera JA, Wiegand TLT, Allmendinger S, Dimitriadis K, Koerte IK. ChatGPT passes German state examination in medicine with picture questions omitted. Dtsch Arztebl Int 2023 May 30;120(21):373-374. [doi: 10.3238/arztebl.m2023.0113] [Medline: 37530052]
41. Gilson A, Safranek CW, Huang T, et al. How does ChatGPT perform on the United States Medical Licensing Examination (USMLE)? The implications of large language models for medical education and knowledge assessment. JMIR Med Educ 2023 Feb 8;9:e45312. [doi: 10.2196/45312] [Medline: 36753318]
42. Choi JH, Hickman KE, Monahan AB, Schwarcz D. ChatGPT goes to law school. J Legal Educ 2021;71:387-400. [doi: 10.2139/ssrn.4335905]
43. Freitas PM, Gomes LM. Does ChatGPT pass the Brazilian bar exam? In: Moniz N, Vale Z, Cascalho J, Silva C, Sebastião R, editors. Progress in Artificial Intelligence: 22nd EPIA Conference on Artificial Intelligence, EPIA 2023, Faial Island, Azores, September 5–8, 2023, Proceedings, Part II: Springer; 2023. [doi: 10.1007/978-3-031-49011-8_11]
44. de Winter JCF. Can ChatGPT pass high school exams on English language comprehension? Int J Artif Intell Educ 2024 Sep;34(3):915-930. [doi: 10.1007/s40593-023-00372-z]
45. Lewandowski M, Łukowicz P, Świetlik D, Barańska-Rybak W. ChatGPT-3.5 and ChatGPT-4 dermatological knowledge level based on the Specialty Certificate Examination in Dermatology. Clin Exp Dermatol 2024 Jun 25;49(7):686-691. [doi: 10.1093/ced/llad255] [Medline: 37540015]
46. Brown KE, Yan C, Li Z, et al. Large language models are less effective at clinical prediction tasks than locally trained machine learning models. J Am Med Inform Assoc 2025 May 1;32(5):811-822. [doi: 10.1093/jamia/ocaf038] [Medline: 40056436]
47. Simon HA. A behavioral model of rational choice. Q J Econ 1955 Feb;69(1):99. [doi: 10.2307/1884852]

48. Hertwig R, Leuker C, Pachur T, Spiliopoulos L, Pleskac TJ. Studies in ecological rationality. Top Cogn Sci 2022 Jul;14(3):467-491. [doi: 10.1111/tops.12567] [Medline: 34310848]
49. Klein G, Jarosz A. A naturalistic study of insight. J Cogn Eng Decis Mak 2011 Dec;5(4):335-351. [doi: 10.1177/1555343411427013]
50. Gigerenzer G, Brighton H. Homo heuristicus: why biased minds make better inferences. Top Cogn Sci 2009 Jan;1(1):107-143. [doi: 10.1111/j.1756-8765.2008.01006.x] [Medline: 25164802]
51. Gigerenzer G, Gaissmaier W. Heuristic decision making. Annu Rev Psychol 2011;62:451-482. [doi: 10.1146/annurev-psych-120709-145346] [Medline: 21126183]
52. Wang Y, Luan S, Gigerenzer G. Modeling fast-and-frugal heuristics. Psych J 2022 Aug;11(4):600-611. [doi: 10.1002/pchj.576] [Medline: 35778774]
53. Artinger FM, Gigerenzer G, Jacobs P. Satisficing: integrating two traditions. J Econ Lit 2022 Jun;60(2):598-635. [doi: 10.1257/jel.20201396]
54. Grove WM, Zald DH, Lebow BS, Snitz BE, Nelson C. Clinical versus mechanical prediction: a meta-analysis. Psychol Assess 2000 Mar;12(1):19-30. [doi: 10.1037/1040-3590.12.1.19] [Medline: 10752360]
55. Karelaia N, Hogarth RM. Determinants of linear judgment: a meta-analysis of lens model studies. Psychol Bull 2008 May;134(3):404-426. [doi: 10.1037/0033-2909.134.3.404] [Medline: 18444703]
56. Klein GA. A recognition-primed decision (RPD) model of rapid decision making. In: Klein GA, Orasanu J, Calderwood R, Zsambok CE, editors. Decision Making in Action: Models and Methods: Ablex Publishing; 1993:138-147 URL: https://psycnet.apa.org/record/1993-97634-006 [accessed 2026-03-18]
57. Klein G, Phillips JK, Rall EL, Peluso DA. A data-frame theory of sensemaking. In: Hoffman RR, editor. Expertise Out of Context: Proceedings of the Sixth International Conference on Naturalistic Decision Making: Lawrence Erlbaum Associates Publishers; 2007:113-155. [doi: 10.4324/9780203810088]
58. Reale C, Salwei ME, Militello LG, et al. Decision-making during high-risk events: a systematic literature review. J Cogn Eng Decis Mak 2023 Jun;17(2):188-212. [doi: 10.1177/15553434221147415] [Medline: 37823061]
59. Ahuna JK, Becker KD. Scoping review of naturalistic decision making studies among mental health professionals: coverage of characteristics and contexts. J Cogn Eng Decis Mak 2025 Mar;19(1):96-129. [doi: 10.1177/15553434241303806]
60. Kopka M, Feufel MA. How to evaluate the accuracy of symptom checkers and diagnostic decision support systems: Symptom Checker Accuracy Reporting Framework (SCARF). JMIR Hum Factors 2026 Jan 16;13:e76168. [doi: 10.2196/76168] [Medline: 41544248]
61. Semigran HL, Linder JA, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self diagnosis and triage: audit study. BMJ 2015 Jul 8;351:h3480. [doi: 10.1136/bmj.h3480] [Medline: 26157077]
62. Hill MG, Sim M, Mills B. The quality of diagnosis and triage advice provided by free online symptom checkers and apps in Australia. Med J Aust 2020 Jun;212(11):514-519. [doi: 10.5694/mja2.50600] [Medline: 32391611]
63. Sætre LMS, Raasthøj I, Lauridsen GB, et al. Revisiting the symptom iceberg based on the Danish symptom cohort - symptom experiences and healthcare-seeking behaviour in the general Danish population in 2022. Heliyon 2024 May 30;10(10):e31090. [doi: 10.1016/j.heliyon.2024.e31090] [Medline: 38803940]
64. McAteer A, Elliott AM, Hannaford PC. Ascertaining the size of the symptom iceberg in a UK-wide community-based survey. Br J Gen Pract 2011 Jan;61(582):e1-e11. [doi: 10.3399/bjgp11X548910] [Medline: 21401979]
65. Kopka M, Scatturin L, Napierala H, et al. Characteristics of users and nonusers of symptom checkers in Germany: cross-sectional survey study. J Med Internet Res 2023 Jun 20;25:e46231. [doi: 10.2196/46231] [Medline: 37338970]
66. Brown HL. Evaluation of artificial intelligence for patient self-triage: comparison of general-purpose AI platforms with the NHS 111 online symptom checker in the United Kingdom. Cureus 2025 Nov;17(11):e97834. [doi: 10.7759/cureus.97834] [Medline: 41458776]
67. Kopka M, He L, Feufel MA. Evaluating the accuracy of ChatGPT model versions for giving care-seeking advice. Commun Med (Lond) 2026 Feb 25;6(1):171. [doi: 10.1038/s43856-026-01466-0] [Medline: 41735505]
68. Fraser H, Crossland D, Bacher I, Ranney M, Madsen T, Hilliard R. Comparison of diagnostic and triage accuracy of Ada Health and WebMD symptom checkers, ChatGPT, and physicians for patients in an emergency department: clinical data analysis study. JMIR mHealth uHealth 2023 Oct 3;11:e49995. [doi: 10.2196/49995] [Medline: 37788063]
69. Levine DM, Tuwani R, Kompa B, et al. The diagnostic and triage accuracy of the GPT-3 artificial intelligence model: an observational study. Lancet Digit Health 2024 Aug;6(8):e555-e561. [doi: 10.1016/S2589-7500(24)00097-9] [Medline: 39059888]
70. Kopka M, Napierala H, Privoznik M, Sapunova D, Zhang S, Feufel MA. The RepVig framework for designing use-case specific representative vignettes and evaluating triage accuracy of laypeople and symptom assessment applications. Sci Rep 2024 Dec 23;14(1):30614. [doi: 10.1038/s41598-024-83844-z] [Medline: 39715767]
71. Gallifant J, Afshar M, Ameen S, et al. The TRIPOD-LLM reporting guideline for studies using large language models. Nat Med 2025 Jan;31(1):60-69. [doi: 10.1038/s41591-024-03425-5] [Medline: 39779929]
72. Fletcher R, Nielsen RK. What does the public in six countries think of generative AI in news? : Reuters Institute for the Study of Journalism, University of Oxford; 2024 URL: https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2024-05/Fletcher_and_Nielsen_Generative_AI_and_News_Audiences.pdf [accessed 2026-03-18] [doi: 10.60625/RISJ-4ZB8-CG87]

73. Schmieding ML, Kopka M, Schmidt K, Schulz-Niethammer S, Balzer F, Feufel MA. Triage accuracy of symptom checker apps: 5-year follow-up evaluation. J Med Internet Res 2022 May 10;24(5):e31810. [doi: 10.2196/31810] [Medline: 35536633]
74. Ceney A, Tolond S, Glowinski A, Marks B, Swift S, Palser T. Accuracy of online symptom checkers and the potential impact on service utilisation. PLoS One 2021;16(7):e0254088. [doi: 10.1371/journal.pone.0254088] [Medline: 34265845]
75. Hirosawa T, Kawamura R, Harada Y, et al. ChatGPT-Generated differential diagnosis lists for complex case-derived clinical vignettes: diagnostic accuracy evaluation. JMIR Med Inform 2023 Oct 9;11:e48808. [doi: 10.2196/48808] [Medline: 37812468]
76. Garg RK, Urs VL, Agarwal AA, Chaudhary SK, Paliwal V, Kar SK. Exploring the role of ChatGPT in patient care (diagnosis and treatment) and medical research: a systematic review. Health Promot Perspect 2023;13(3):183-191. [doi: 10.34172/hpp.2023.22] [Medline: 37808939]
77. Hose BZ, Rounds AK, Nandwani I, et al. Use of ChatGPT for urinary symptom management among people with spinal cord injury or disease: qualitative study. JMIR Rehabil Assist Technol 2025 May 29;12:e70339. [doi: 10.2196/70339] [Medline: 40440564]
78. Al Shboul MKI, Alwreikat A, Alotaibi FA. Investigating the use of ChatGPT as a novel method for seeking health information: a qualitative approach. Sci Technol Libr 2024 Jul 2;43(3):225-234. [doi: 10.1080/0194262X.2023.2250835]
79. Ayo-Ajibola O, Davis RJ, Lin ME, Riddell J, Kravitz RL. Characterizing the adoption and experiences of users of artificial intelligence-generated health information in the United States: cross-sectional questionnaire study. J Med Internet Res 2024 Aug 14;26:e55138. [doi: 10.2196/55138] [Medline: 39141910]
80. Liu J, Wang C, Liu S. Utility of ChatGPT in clinical practice. J Med Internet Res 2023 Jun 28;25:e48568. [doi: 10.2196/48568] [Medline: 37379067]
81. Kopka M, Feufel MA. Statistical refinement of patient-centered case vignettes for digital health research. Front Digit Health 2024;6:1411924. [doi: 10.3389/fdgth.2024.1411924] [Medline: 39498100]
82. Kopka M, Feufel MA, Berner ES, Schmieding ML. How suitable are clinical vignettes for the evaluation of symptom checker apps? A test theoretical perspective. Digit Health 2023;9:20552076231194929. [doi: 10.1177/20552076231194929] [Medline: 37614591]
83. Painter A, Hayhoe B, Riboli-Sasco E, El-Osta A. Online symptom checkers: recommendations for a vignette-based clinical evaluation standard. J Med Internet Res 2022 Oct 26;24(10):e37408. [doi: 10.2196/37408] [Medline: 36287594]
84. Arellano Carmona K, Chittamuru D, Kravitz RL, Ramondt S, Ramírez AS. Health information seeking from an intelligent web-based symptom checker: cross-sectional questionnaire study. J Med Internet Res 2022 Aug 19;24(8):e36322. [doi: 10.2196/36322] [Medline: 35984690]
85. Ito N, Kadomatsu S, Fujisawa M, et al. The accuracy and potential racial and ethnic biases of GPT-4 in the diagnosis and triage of health conditions: evaluation study. JMIR Med Educ 2023 Nov 2;9:e47532. [doi: 10.2196/47532] [Medline: 37917120]
86. Epstein RH, Dexter F. Variability in large language models' responses to medical licensing and certification examinations. Comment on "How does ChatGPT perform on the United States Medical Licensing Examination? The implications of large language models for medical education and knowledge assessment". JMIR Med Educ 2023 Jul 13;9:e48305. [doi: 10.2196/48305] [Medline: 37440293]
87. Franc JM, Cheng L, Hart A, Hata R, Hertelendy A. Repeatability, reproducibility, and diagnostic accuracy of a commercial large language model (ChatGPT) to perform emergency department triage using the Canadian triage and acuity scale. CJEM 2024 Jan;26(1):40-46. [doi: 10.1007/s43678-023-00616-w] [Medline: 38206515]
88. Kopka M, Feufel MA. Software symptomcheckR: an R package for analyzing and visualizing symptom checker triage performance. BMC Digit Health 2024 Jul 22;2(1):43. [doi: 10.1186/s44247-024-00096-7]
89. Wickham H, Averick M, Bryan J, et al. Welcome to the Tidyverse. J Open Source Softw 2019 Nov 21;4(43):1686. [doi: 10.21105/joss.01686]
90. Psych: procedures for psychological, psychometric, and personality research. The Comprehensive R Archive Network. URL: https://CRAN.R-project.org/package=psych [accessed 2025-08-16]
91. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Soft 2015;67(1). [doi: 10.18637/jss.v067.i01]
92. Jędrzejczak WW, Skarżyński H, Kochanek K. Testing new versions of ChatGPT in terms of physiology and electrophysiology of hearing: improved accuracy but not consistency. medRxiv. Preprint posted online on Oct 8, 2024. [doi: 10.1101/2024.10.08.24315089]
93. Flaharty KA, Hu P, Hanchard SL, et al. Evaluating large language models on medical, lay-language, and self-reported descriptions of genetic conditions. Am J Hum Genet 2024 Sep 5;111(9):1819-1833. [doi: 10.1016/j.ajhg.2024.07.011] [Medline: 39146935]
94. Moore J, Grabb D, Agnew W, et al. Expressing stigma and inappropriate responses prevents LLMs from safely replacing mental health providers. 2025 Jun 23 Presented at: FAccT ’25; Jun 23-26, 2025 URL: https://dl.acm.org/doi/proceedings/10.1145/3715275 [doi: 10.1145/3715275.3732039]

95. Omar M, Sorin V, Collins JD, et al. Multi-model assurance analysis showing large language models are highly vulnerable to adversarial hallucination attacks during clinical decision support. Commun Med (Lond) 2025 Aug 2;5(1):330. [doi: 10.1038/s43856-025-01021-3] [Medline: 40753316]
96. Hollnagel E, Pedersen OM, Rasmussen J. Notes on human performance analysis. : Risø National Laboratory; 1981 URL: https://backend.orbit.dtu.dk/ws/portalfiles/portal/88561089/ris_m_2285.pdf [accessed 2026-03-18]
97. Zhang S, Yu J, Xu X, et al. Rethinking human-AI collaboration in complex medical decision making: a case study in sepsis diagnosis. Presented at: Proceedings of the CHI Conference on Human Factors in Computing Systems; May 11-16, 2024. [doi: 10.1145/3613904.3642343]
98. Kämmer JE, Ernst K, Grab K, et al. Collaboration during the diagnostic decision‐making process: when does it help? J Behav Decis Mak 2024 Jan;37(1):e2357. [doi: 10.1002/bdm.2357]
99. Kopka M. Data set supplementing “increasing large language model accuracy for care-seeking advice using prompts reflecting human reasoning strategies in the real world: validation study”. Zenodo. 2026 Mar 28. URL: https://zenodo.org/records/19291719 [accessed 2026-04-06]

Abbreviations

AI: artificial intelligence
HF/E: human factors and ergonomics
LLM: large language model
NDM: naturalistic decision-making
RPD: recognition-primed decision-making
TRIPOD: Transparent Reporting of a Multivariable Model for Individual Prognosis or Diagnosis

Edited by J Sarvestan; submitted 18.Nov.2025; peer-reviewed by C Baxter; revised version received 04.Mar.2026; accepted 04.Mar.2026; published 08.Apr.2026.

Please cite as:
Kopka M, Feufel MA
Increasing Large Language Model Accuracy for Care-Seeking Advice Using Prompts Reflecting Human Reasoning Strategies in the Real World: Validation Study
JMIR Biomed Eng 2026;11:e88053
URL: https://biomedeng.jmir.org/2026/1/e88053
doi:10.2196/88053

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org