

# PAIGE Chatbot For Patient-Clinician Communication: Usability and Utility Assessment

Agency for Healthcare Research and Quality  
5600 Fishers Lane  
Rockville, MD 20857  
[www.ahrq.gov](http://www.ahrq.gov)

Contract No: 75Q80120D00018

**Prepared by:**

Courtney Zott, MPH  
Dean F. Sittig, PhD  
Nikki Gauthreaux, MPH  
Adam Wright, PhD, FACMI, FAMIA, FIAHSI  
Elise Russo, MPH, PMP  
Laura Zahn, MS  
Prashila M. Dullabh, MD

August 2024



---

## PURPOSE

The Clinical Decision Support Innovation Collaborative (CDSiC) aims to advance the design, development, dissemination, implementation, use, measurement, and evaluation of evidence-based, shareable, interoperable, and publicly available patient-centered clinical decision support (PC CDS) to improve health outcomes of all patients by creating a proving ground of innovation. The Innovation Center supports the measurement of PC CDS implementation and effectiveness to ensure that PC CDS works as intended. This report is intended for those interested in developing, implementing, and evaluating AI-based PC CDS technology. All qualitative research activities conducted by the CDSiC are reviewed by the NORC at the University of Chicago Institutional Review Board (FWA00000142).

---

## FUNDING STATEMENT

This project was funded under contract number 75Q80120D00018 from the Agency for Healthcare Research and Quality (AHRQ), U.S. Department of Health and Human Services (HHS). The opinions expressed in this document are those of the authors and do not reflect the official position of AHRQ or HHS.

---

## PUBLIC DOMAIN NOTICE

This document is in the public domain and may be used and reprinted without permission in the United States unless materials are clearly noted as copyrighted in the document. No one may reproduce copyrighted materials without the permission of the copyright holders. Users outside the United States must get permission from AHRQ to reprint or translate this product. Anyone wanting to reproduce this product for sale must contact AHRQ for permission. Citation of the source is appreciated.

---

## SUGGESTED CITATION

Zott C, Sittig DF, Gauthreaux N, Wright A, Russo E, Zahn L, Dullabh PM.. PAIGE Chatbot for Patient-Clinician Communication: Usability and Utility Assessment. Prepared under Contract No. 75Q80120D00018. Rockville, MD: Agency for Healthcare Research and Quality; August 2024.

## Acknowledgements

NORC would like to extend its sincere thanks to Dr. Edwin Lomotan and James Swiger at AHRQ for their strong leadership and steadfast support for this project. NORC would like to thank the technical team at Vanderbilt University Medical Center that contributed to the development of PAIGE. In addition, NORC would also like to thank the clinicians and patient representatives who dedicated time to provide us with valuable insights and feedback. Finally, NORC would like to express their deep gratitude for the time and expertise provided by the Innovation Center Planning Committee, without whom this work would not have been possible.

# Table of Contents

- Executive Summary ..... 1
  - Introduction ..... 1
  - Purpose ..... 1
  - Methods ..... 1
  - Results ..... 2
  - Discussion ..... 3
  - Conclusion ..... 4
- 1. Introduction ..... 5
  - 1.1. Background and Relevance to Research ..... 5
- 2. Technical Overview of PAIGE Chatbot ..... 6
- 3. Assessment Overview ..... 9
  - 3.1 Research Questions ..... 9
  - 3.2 Data Collection Methods ..... 10
    - 3.2.1 Patient Usability Assessment ..... 10
    - 3.2.2 Clinician Utility Assessment ..... 11
  - 3.3 Data Analysis ..... 11
    - 3.3.1 Qualitative Analysis ..... 11
    - 3.3.2 Quantitative Analysis ..... 12
    - 3.3.3 Mixed-Methods Analysis ..... 12
- 4. Assessment Results ..... 12
  - 4.1 Assessment Goal 1: Patients’ Ability to Comprehend and Respond to PAIGE ..... 12
    - 4.1.1 PAIGE and Patient Interaction Details ..... 13
    - 4.1.2 Modified eHealth Literacy Scale Survey Results ..... 13
    - 4.1.3 Readability Results ..... 14
    - 4.1.4 Key Qualitative Themes ..... 15
  - 4.2 Assessment Goal 2: Quality and Utility of PAIGE-Generated Summaries ..... 17
    - 4.2.1 Suggestions to Improve PAIGE Queries ..... 18
    - 4.2.2 Suggestions for PAIGE-Generated Summaries ..... 19
    - 4.2.3 Changes to Health Care Delivery ..... 19
    - 4.2.4 Discussion ..... 19
- 5. Conclusion ..... 20

Appendix A: Example Interaction Between a Patient and PAIGE .....	21
Appendix B. Modified eHealth Literacy Scale .....	22
Appendix C. Instrument for Clinician Assessment .....	23
References.....	25

# Executive Summary

## Introduction

Over the past two decades, patient demand for electronic communication with their clinicians has increased steadily, facilitated in part by increased use of patient portals with messaging capabilities. One frustration with portal communication for patients and clinicians alike is the time and effort to collect and synthesize patient-reported information for the clinician to respond effectively. Conversational artificial intelligence (AI) chatbots exist in many business settings, deployed as human-like resources answering questions for people needing customer support. Large language model (LLM)-based tools like ChatGPT are increasingly used in health care applications and can potentially reduce back-and-forth messaging to respond more efficiently to patient questions.

The Agency for Healthcare Research and Quality (AHRQ) Clinical Decision Support Innovation Collaborative (CDSiC) Innovation Center developed a prototype patient-centered clinical decision support (PC CDS) tool, the Patient Artificial Intelligence-Guided E-messages (PAIGE) chatbot. PAIGE leverages an LLM to improve and streamline patient-clinician communication. This report presents findings from an assessment of the usability and utility of the PAIGE prototype chatbot, providing insights into key design and usage considerations, as well as identifying issues for consideration by developers of patient-facing AI-based chatbots.

## Purpose

The assessment aimed to: 1) assess patients' ability to comprehend and respond to PAIGE questions and 2) assess the quality and utility of PAIGE-generated summaries for clinicians. The overarching aim was to better understand the capability of LLMs to facilitate accurate and timely clinical communication between patients and clinicians, with an ultimate goal of identifying opportunities and challenges associated with this new class of patient-facing, CDS functionality.

## Methods

We conducted five "think-aloud" usability testing sessions with paid patient representatives identified by CDSiC members. Patient participants interacted with PAIGE using hypothetical clinical questions. Following the session, we surveyed patients about using PAIGE to send messages to their clinician. The survey asked for participants' agreement with statements about PAIGE using a 5-point Likert scale, where 1 meant strongly disagree with the statement and 5 meant strongly agree. We then selected the 10 most informative, useful, or complete set of user interactions and calculated quantitative metrics related to the timing of responses and reading level. Finally, we asked six clinicians to independently rate their agreement regarding the appropriateness, completeness, utility, accuracy, and other aspects of the 10 PAIGE-patient interactions, using the same Likert scale. We also encouraged the clinicians to provide us with other feedback about these interactions. We used mixed methods to analyze and synthesize the key themes from this data.

## Results

**Assessment Goal 1: Patients' Ability to Comprehend and Respond to PAIGE:** On average, PAIGE responded to patient messages and generated summaries quickly (6 and 15 seconds, respectively), at a middle-school reading level. Generally, patients strongly agreed that PAIGE was helpful in formulating a question about their health and disagreed strongly that PAIGE used words that were difficult to understand. Qualitatively, patients provided the following feedback regarding what they liked about PAIGE, what could be improved, what features could be added, and what activities could refine the prototype in the future:

- **What Patients Liked**
  - PAIGE's summaries were accurate, appropriate, and comprehensive.
  - PAIGE appropriately used AI by collecting patient information but not providing medical advice.
  - PAIGE reduced the cognitive burden on patients to know what information and symptoms to provide to their clinician.
- **Opportunities for Improvement**
  - PAIGE could feel like an "intermediary" that lessens patients' personal connections with their clinicians.
  - PAIGE may not be all patients' preferred way to communicate with clinicians.
  - While PAIGE reduced cognitive burden, it may not reduce time to communicate with clinicians.
  - A 48-hour timeframe for a response from clinicians is too long and may deter use.
- **Suggestions for Refinements and Future Work on Similar Patient-Facing Chatbots**
  - AI-generated summaries could better incorporate the patient voice by retaining language patients used to describe their symptoms, versus translating their colloquial terms into formal clinical language.
  - AI-generated summaries could provide additional links to or include vetted clinical information or resources for patients to read while they are waiting for their clinician's response.
  - AI-generated interactions could incorporate email functionality to enable users to email the AI-generated summary to themselves and/or to a family member or caregiver to provide a record of their communication.
  - More usability testing should be conducted with patients with limited digital and/or health literacy.
  - AI-generated summaries should include access to information from the electronic health record (EHR) to avoid asking patients questions about information the clinician and/or health care system should already know (e.g., current medications, treatment history).

**Assessment Goal 2: Quality and Utility of PAIGE-Generated Summaries:** Clinicians strongly agreed that PAIGE summaries accurately captured the patient's original questions and responses.

They agreed least with statements that PAIGE's clarifying questions were complete and no other information would be required to respond to the patient's original question. Most clinicians reported they preferred the PAIGE-generated summary to the patient's original question. Qualitatively, clinicians reported that PAIGE showed promise as a PC CDS tool with improvements needed to the questions asked and details included in the summaries, as well as to the broader health care delivery landscape.

- **Suggestions to Improve AI-Generated Questions:** Clinicians reported that in some scenarios AI should ask additional questions about treatment history to elucidate information that may be unavailable in the EHR. This would better facilitate decision-making about next steps in the patient's care. However, clinicians also noted that AI should adhere to standard triage questions for common conditions (e.g., upper respiratory infection) even when medical or treatment history is known, as recent changes in the patient's treatment or other co-occurring clinical conditions are the most likely causes. Finally, clinicians noted that AI could ask more relevant follow-up questions if it had more patient context. For example, age is critical to conducting differential diagnosis as it impacts the individual's probability for a particular condition or disease (e.g., menopause).
- **Suggestions for AI-Generated Summaries:** Clinicians suggested adding relevant patient background information from the EHR (e.g., age). They also noted that the summaries could be improved by including information when a patient explicitly denies having a symptom (e.g. no fever, no chills, no shortness of breath).
- **Changes to Health Care Delivery:** Clinicians noted that fee-for-service is still the predominate payment method, and their time to interact with patients via portal messages is limited (versus compensated in-person visits). They noted that other types of health care delivery models are better at incentivizing time spent on portal messages, such as value-based care models and scheduled e-visits. Without some acknowledgement that these asynchronous patient interactions can consume a nontrivial portion of a clinician's work day (and night), AI-based messaging systems, while potentially of benefit to patients, may contribute to clinician overwork and burnout.

## Discussion

Our findings indicate general patient and clinician enthusiasm for PAIGE and more generally the concept of AI-generated patient-clinician communication. Despite highly relevant recommendations for improvements and suggestions for future testing, patients and clinicians reported that AI-generated summaries held promise for enhancing patient-clinician communication for many common clinical concerns and questions. We found that patients thought PAIGE was easy to understand and clinicians thought PAIGE was useful as a tool for summarizing patients' clinical concerns. Future prototypes and testing should focus on ensuring AI-generated summaries save patients and clinicians time, integrating contextual information from the EHR about patients, and adding more patient-centered features. The generally positive reception of PAIGE and AI-generated summaries in general among patients and clinicians is an optimistic sign for developers of similar tools. However, the suggestions for improvements and future work reveal important considerations for the design and success of such



tools. Our assessment begins to provide insight into the specific features of AI-based chatbots that will enable them to save time and reduce cognitive burden for both clinicians and patients, as well as feel like a service to patients rather than an extra step in the communication process. However, more research is necessary to understand whether and how AI-based chatbots can streamline communication between clinicians and patients, and which clinician types and patient populations will benefit most from such tools.

## Conclusion

Conversational AI chatbots have the potential to support and streamline communication between patients and clinicians, but such technologies must be designed carefully with patients and clinicians in mind. As an early prototype of such a chatbot, PAIGE shows promise as an appropriate, usable, and useful AI-based tool for both patient and clinicians end-users. Developers will continue to improve PAIGE and similar AI-generated patient summary tools based on the feedback in this assessment to enhance and refine their products.

# 1. Introduction

NORC at the University of Chicago (NORC) is pleased to submit this final assessment report for the Patient Artificial Intelligence-Guided E-messages (PAIGE) chatbot prototype usability and utility assessment to the Agency for Healthcare Research and Quality (AHRQ). Since 2021, AHRQ has supported the Clinical Decision Support Innovation Collaborative (CDSiC) to advance evidence into practice for patient-centered clinical decision support (PC CDS). PC CDS encompasses a spectrum of decision-making tools that significantly incorporate patient-centered factors related to knowledge, data, delivery, and use. Knowledge refers to the use of comparative effectiveness research or patient-centered outcomes research findings. Data focuses on the incorporation of patient-generated health data, patient preferences, social determinants of health, and other patient-specific information. Delivery refers to directly engaging patients and/or caregivers across different settings. Finally, use focuses on facilitating bidirectional information exchange in support of patient-centered care, including shared decision-making.

This report presents findings from the assessment of the design and development of a PC CDS tool, the PAIGE chatbot, that leverages a large language model (LLM) to improve patient-clinician communication. In addition to AHRQ, the report findings may be of interest to researchers, informaticians, developers, hospital administrators, clinicians, policymakers, and government agencies seeking to develop and implement PC CDS.

## 1.1. Background and Relevance to Research

Successful communication between patients and clinicians is essential for high-quality, patient-centered care. Over the past two decades, patient demand for electronic communication with their clinicians has steadily increased, facilitated in part by increased patient portal use.<sup>1</sup> A cross-sectional survey of patients with access to a patient portal for clinician messaging determined those with higher out-of-pocket costs for in-person visits ( $\geq \$60$ ) were significantly more likely to choose electronic messaging as their first method of contact, and 32% of participants reported that being able to communicate via messaging improved their overall health.<sup>2</sup> However, because electronic messages are asynchronous and can occur outside business hours, clinicians have indicated challenges with receiving and responding to large volumes of patient messages.<sup>3</sup> Studies have estimated that primary care clinicians spend an average of 1.5 hours a day processing inbox messages related to patient care,<sup>4</sup> with approximately 150 inbox messages a day.<sup>5</sup>

One source of frustration with portal communication for patients and clinicians alike is the time and effort needed to gather enough information from the patient for the clinician to respond effectively. A common example occurs when a patient requests prescriptions for nirmatrelvir and ritonavir (Paxlovid). Patients may write, "I have a positive COVID test, can you prescribe Paxlovid," but the clinician needs additional information to determine whether the prescription is appropriate (e.g., positive test date, risk factors, and additional current medications). However, patients are only eligible for Paxlovid within five days of a positive test, leading to situations where patients can time out of eligibility for therapy during prolonged message exchanges with clinicians. A tool that helps patients compose more complete and

effective messages to clinicians can potentially improve communication, decision-making, and satisfaction for both patients and clinicians.

Conversational artificial intelligence (AI) chatbots exist in many business settings, deployed as human-like resources answering questions for people needing customer support. Despite their prevalence, a recent survey of 1,000 U.S. consumers determined 86% of people would prefer to talk to a human instead of engaging with an AI-based system, in large part due to the belief that virtual assistants make it harder to get their issues resolved.<sup>6</sup> In recent years, LLM-based tools like ChatGPT increasingly have been used in health care applications, namely in the areas of clinical decision support, medical documentation, education, and contextual-question answering.<sup>7,8,9</sup> The evidence reported has been mostly positive in terms of effectiveness and user satisfaction, with evaluators indicating responses from chatbots to be significantly more empathetic and higher quality compared to those from physicians.<sup>10,11</sup> Given the volume of back-and-forth messaging that can occur between patients and providers, there is an opportunity to use LLM technology to better support communication. Rather than generating automated answers to patient questions directly—a topic already widely researched and available in some commercial electronic health records (EHRs)—LLM technology can be used to clarify patient needs and symptoms from the beginning, enabling providers to respond to messages more efficiently and effectively.

## 2. Technical Overview of PAIGE Chatbot

As a tool that helps patients compose more effective messages to their clinicians, PAIGE has the potential to improve the understanding and efficiency of patient and clinician communication, decision-making, and satisfaction. The system takes as input a patient's question and generates one or more clarifying questions that are fed back to the patient sequentially to simulate a back-and-forth conversation. The questions and responses to patients do not provide medical advice. When the patient has answered all PAIGE questions, PAIGE generates a summary of the patient's original question along with their responses to the follow-up clarifying questions. This summary can then be reviewed by the patient and, if deemed appropriate, sent to their clinician for an answer. Eventually, the system is intended to be integrated into a health system patient portal and leverage patients' EHR data to tailor follow-up questions. However, due to the limited scope of this project and timeline, the system was developed as a standalone, web-based simulation that does not use real patient data from any existing clinical information system. In this section, we describe the technical components of the chatbot and the co-design process.

Currently, the PAIGE chatbot prototype exists in two forms: 1) a simulated, static patient portal with pre-loaded conversation scenarios (see Exhibit 1) and 2) an interactive ChatGPT environment (see Exhibit 2).

**Exhibit 1.** A mock-up showing how PAIGE might look if it were integrated into a patient portal. The green boxes illustrate the patient's questions and answers, while the white boxes illustrate the questions generated by PAIGE. The ultimate vision is for PAIGE to be integrated into the health system patient portal and EHR.

The mock-up displays a patient portal interface. On the left, there are three main sections: **Patient Information** (Name: Jane Doe, DOB: 01/01/1980, Last Visit: 03/15/2023), **Upcoming Appointments** (04/10/2023 - Regular Checkup, 05/12/2023 - Follow-up Consultation), and **View Results** (Blood Test Results - 03/15/2023, X-Ray Report - 02/10/2023). The right side features a **Message Your Provider** chat interface. At the top of the chat are four buttons: **UTI (ChatGPT)**, **UTI (CLAIR)**, **Example 3**, and **Example 4**. The chat history shows a patient message: "I have a bad UTI. Can you send me in a prescription?". A white box contains a disclaimer: "Your message will be shared with your physician for response. We use AI to review patient messages and identify potential follow-up questions or clarifications. This allows your physician to respond more quickly and completely. Please consider replying with this information:". Below this, a white box asks: "Is there a strong urge to urinate even when your bladder is not full?". A green box replies: "Yes". Another white box asks: "AI: Do you have pain or a burning sensation when you urinate?". A green box replies: "Not really, maybe a little bit. Mostly just the urge to pee that won't go away.". A third white box asks: "AI: Have you noticed any blood in your urine?". A green box replies: "Yes I think so, but only a little bit. It's very light.". A final white box asks: "AI: Do you feel any lower abdominal pain or discomfort?".

**Exhibit 2.** Interactive version of PAIGE that was used for the assessment in this report. In this illustration, the lines that begin with “You” show the patient’s question and subsequent responses to PAIGE-generated queries, while the lines that begin with “PAIGE” show PAIGE’s responses to the patient’s question.

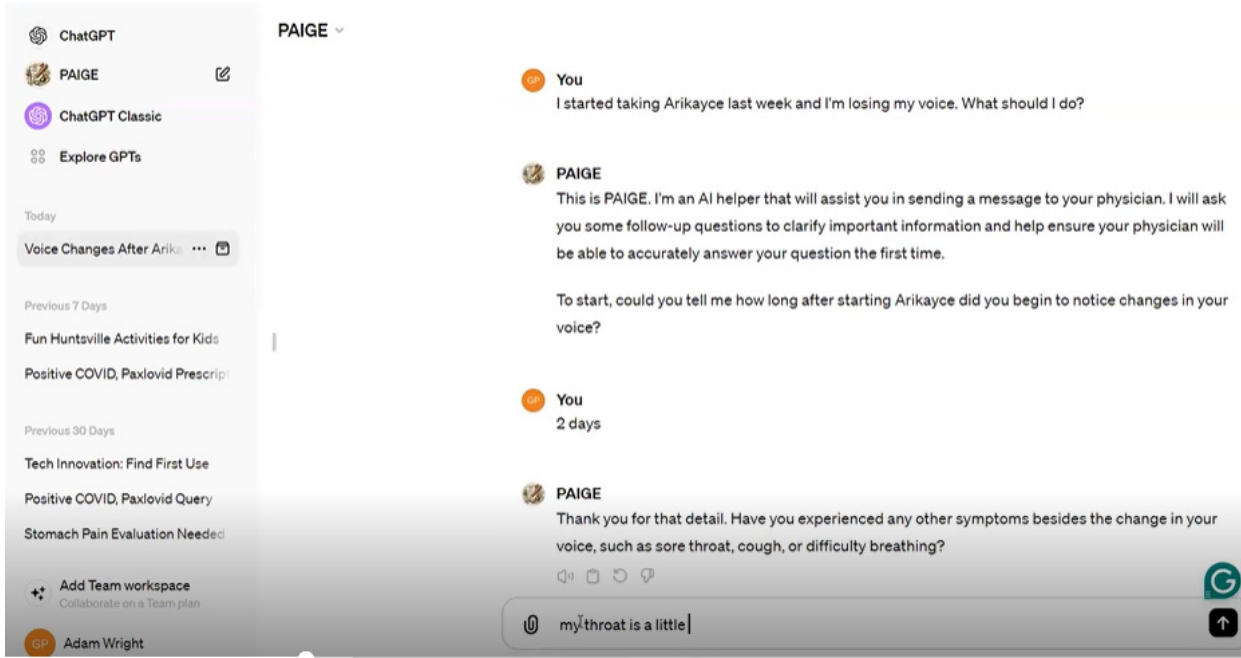
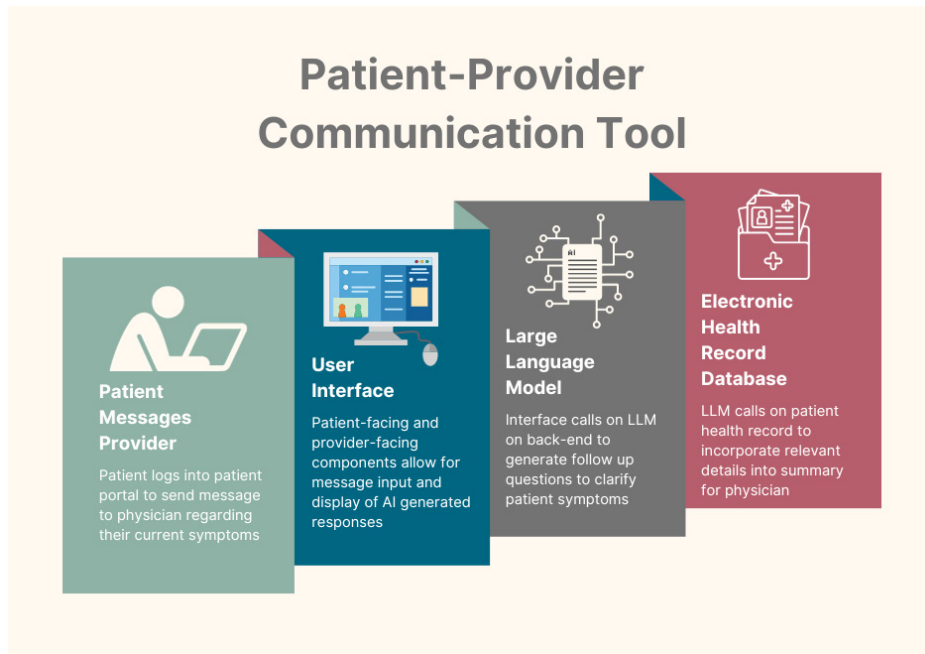


Exhibit 3 highlights the technical components of PAIGE, including the user interface and underlying LLM components. PAIGE would have both a patient-facing and clinician-facing user interface. On the patient side, the interface would exist within the patient portal and leverage ChatGPT LLM technology to receive and respond to patient messages or requests to their clinicians with clarifying questions. On the clinician side, the interface would be integrated within the EHR’s clinician inbox and would leverage ChatGPT LLM technology to display a summary of the patient’s responses. If possible and where appropriate, the PAIGE LLM would extract relevant patient information from the EHR database and use that information to increase the specificity of follow-up questions or the final summary for the clinician.

Once the technology is developed, then any EHR vendor could create a simple, interactive user interface that collects a patient’s question and contacts PAIGE, or another Health Insurance Portability and Accountability Act (HIPAA)-certified, AI-generated patient question and summarization service available via the internet. The AI-generated response is then presented to the patient user who responds to the AI-generated questions until the interaction is complete. Then the AI-generated summary of the patient’s initial question and responses is summarized and shown to the patient for confirmation or editing. Once the summarization is complete, the AI application generates a message to the clinician’s in-box, and the clinician reads the message, much as they would read any other message generated by a patient. The clinician then responds to this message, just as they would to any human-generated message, and the patient receives the clinician response in their patient portal.

**Exhibit 3.** Overview of the Patient-Clinician Communication System Components



### 3. Assessment Overview

NORC conducted a formative, mixed-methods assessment of the PAIGE chatbot between April and July 2024. The assessment focused on one aspect of the usability of PAIGE—specifically, PAIGE’s ability to generate clear and understandable responses in the form of questions for patients and a useful summary of information for clinician decision-making. The assessment did not focus on the look and feel of the PAIGE user interface, as it was not yet integrated into the patient portal. The assessment’s overarching aim was to generate a better understanding of the capability of LLMs to facilitate clinical communication between patients and clinicians, with the goal of improving future design and development of similar interventions.

Specifically, the goals of the formative assessment were to:

- **Goal 1:** Assess patients’ ability to comprehend and respond to PAIGE questions.
- **Goal 2:** Assess the quality and utility of PAIGE-generated summaries for clinicians.

In this section, we describe the research questions aligned with these goals as well as the data collection and analysis methods for the assessment.

#### 3.1 Research Questions

To achieve the overarching goals of the assessment, we identified primary research questions specific to each goal (Exhibit 4). In answering these questions, we describe the usability of information generated by the PAIGE system for patient and clinician users.

**Exhibit 4. Key Research Questions for PAIGE Assessment**


Evaluation Goal	Key Research Questions
<b>Assess patients' ability to comprehend and respond to PAIGE questions</b>	<ul style="list-style-type: none"> <li>• To what extent did patients find PAIGE easy to understand?</li> <li>• What types of clarifying questions did patients find difficult to understand? Why?</li> <li>• What types of clarifying questions did patients find irrelevant? Why?</li> <li>• How much time did patients spend communicating with PAIGE?</li> </ul>
<b>Assess the utility and quality of PAIGE-generated summaries for clinicians</b>	<ul style="list-style-type: none"> <li>• To what extent did clinicians find the summaries generated by PAIGE to be useful?</li> <li>• What are clinicians' perceptions of the quality of the summaries generated by PAIGE?</li> </ul>

### 3.2 Data Collection Methods

The assessment collected qualitative and quantitative data from two end-user groups: patients and clinicians. Data sources included “talk- aloud” usability testing sessions, a validated survey instrument, and data from the PAIGE prototype system. All methods and processes for this assessment were reviewed and approved by the NORC Institutional Review Board (IRB) and classified as exempt. NORC is the IRB of record for Vanderbilt University Medical Center.


Exhibit 5 outlines the data collection methods used at each phase of the PAIGE assessment and are further described below.

**Exhibit 5. Evaluation Methods Overview**



**Patient Usability Assessment**

- Conduct 5-7 usability testing sessions with patients
- Administer validated survey instrument to assess comprehension of PAIGE questions
- Collect and analyze inputs and outputs of patient interactions with PAIGE during testing sessions



**Clinician Utility Assessment**

- Administer validated instrument to 3-4 clinicians for rating of usefulness and quality of PAIGE summaries

#### 3.2.1 Patient Usability Assessment

The assessment began with two data collection efforts: 1) qualitative and quantitative data collection from usability testing sessions and 2) quantitative data collection from the prototype PAIGE system.

### 3.2.1.1 Usability Testing Sessions

We conducted five, 45-minute, remote talk-aloud sessions with paid patient representatives identified by CDSiC members.<sup>12</sup> To help maintain patient privacy, each patient was asked to enter a hypothetical question they might send to their clinician. Patients were prompted to add nuance and complexity to their questions, if needed. Patients were then asked to respond to each of PAIGE's clarifying questions. Once PAIGE had gathered enough information, it generated a short summary of the patient's original question along with the patient's responses to the clarifying questions. Each patient was asked to review this summary and whether they would like to add anything to it. Each patient repeated this set of tasks 3-4 times. Following the session, we used a modified version of the eHealth Literacy Scale<sup>13</sup> to gather the patient's opinions on using PAIGE to send messages to their clinician (see Appendix B).

### 3.2.1.2 Post-Session Measures

For each patient, we collected the time required for: a) the patient to enter their initial questions; b) PAIGE to generate the clarifying questions; c) patient to respond to each question; and d) how long it took PAIGE to generate the summary. We also counted the number of words in the user's question and responses to each question, the number of words in PAIGE's clarifying questions and summary, and the Flesch-Kincaid Grade Level metric<sup>14</sup> of the patient's initial question and responses, along with that of the questions and summary from PAIGE. The Flesch-Kincaid Grade Level score takes sentence length and word complexity (e.g., number of syllables) into consideration to determine the approximate reading grade level of text. Scores range from 0-18 corresponding with U.S. grade levels from kindergarten/elementary to the post-graduate level. Lower scores indicate reading levels at lower grades.

NORC offered a \$50 honorarium to all participants. The sessions were conducted via Zoom and lasted approximately 60 minutes.

### 3.2.2 Clinician Utility Assessment

Upon completion of the usability testing sessions, the team selected the 10 most informative, useful, or complete set of user interactions for clinician feedback (based on research team consensus regarding content complexity, challenging nature of the questions, or nuances involved). The team asked six clinicians to independently rate their agreement regarding the appropriateness, completeness, utility, accuracy, and other aspects of the interaction (see instrument in Appendix C).

## 3.3 Data Analysis

The assessment included an analysis of qualitative and quantitative data from each phase of data collection.

### 3.3.1 Qualitative Analysis

**Patient Interactions.** For the usability testing sessions, we conducted a concurrent, rapid thematic analysis to identify challenges and facilitators to patients' understanding of questions from PAIGE. A



research team member coded the transcript-style notes after each session to identify positive and negative feedback, as well as suggestions for improvement. Two research team members then collaborated on reviewing the codes and notes and synthesizing key themes across the sessions. A senior research team member reviewed the key themes to ensure they were comprehensive, useful, and accurate representations of the sessions.

**Clinician Assessment.** For the clinician assessment, two research team members reviewed free-text survey responses and email feedback provided by clinicians and generated key themes. Senior research team members reviewed the themes to ensure they were comprehensive, useful, and accurate representations of the written feedback.

### 3.3.2 Quantitative Analysis

The quantitative analysis consisted of producing univariate descriptive statistics for each phase of the assessment.

**Patient Interactions.** For the usability testing sessions, we calculated statistics related to the study participants' survey scores, time interacting with PAIGE, numbers of words in interactions, and reading level of interaction.

**Clinician Assessment.** For the clinician utility assessment, we calculated statistics on clinicians' ratings of PAIGE's questions and summaries.

### 3.3.3 Mixed-Methods Analysis

Our approach to the mixed-methods analysis was convergent in that we collected qualitative and quantitative data independently but analyzed them concurrently. We integrated qualitative and quantitative data to describe key findings as they relate to the usability and utility of PAIGE for patients and clinicians.

## 4. Assessment Results

This section describes the key findings related to the two assessment goals. The findings are organized by key themes within each assessment goal and provide context for patient and clinician perspectives on PAIGE.

### 4.1 Assessment Goal 1: Patients' Ability to Comprehend and Respond to PAIGE

In this section, we describe the findings of our assessment related to patients' ability to comprehend and respond to PAIGE. We begin with the post-session measures related to time, details, and reading level of PAIGE interactions during the usability testing sessions, followed by qualitative key themes from the sessions.

### 4.1.1 PAIGE and Patient Interaction Details

Exhibit 6 lists details of the 10 interactions selected for the clinician assessment, including time (in seconds) for PAIGE and patient responses and number of clarifying questions asked by PAIGE. For some interactions, the moderator typed for the patient, so patients’ time to initiate and respond to questions could not be calculated. On average, PAIGE took approximately 6 seconds to respond to patient messages and approximately 15 seconds to generate summaries. PAIGE asked approximately four clarifying questions per interaction, and patients took an average of 20 seconds to respond to these questions.

**Exhibit 6.** Details of Patient Interactions with PAIGE

Question	Initial Patient Question (seconds)	# of clarifying questions	PAIGE generates questions (seconds)	Patient responds to questions (seconds)	PAIGE generates summary (seconds)
1	NA	5	8, 4, 3, 7, 5	NA	16
2	NA	4	7, 6, 4, 4	NA	18
3	76	4	12, 3, 8, 5	14, 2, 6, 18	13
4	50	6	11, 3, 8, 9, 5, 4	2, 1, 10, 1, 1, 8	14
5	57	2	6, 6	26, 39	17
6	29	5	10, 5, 6, 4, 5	130, 27, 24, 20, 51	15
7	NA	2	8, 15	NA	13
8	NA	5	13, 3, 6, 4, 4	NA	11
9	34	6	7, 5, 3, 3, 4, 4	6, 17, 12, 8, 44, 28	16
10	30	2	7,6	6, 6	18
<b>Mean</b>	<b>46.0</b>	<b>4.1</b>	<b>6.1</b>	<b>20.3</b>	<b>15.1</b>

### 4.1.2 Modified eHealth Literacy Scale Survey Results

Exhibit 7 details the results for each question of the modified eHealth Literacy Scale survey. Patients could choose from a Likert scale where the strength of agreement increased from 1 to 5 (i.e., 1 meant strongly disagree with the statement and 5 meant strongly agree), and a score of 3 meant the clinician was neutral, or neither agreed nor disagreed. Patients gave the highest mean score to (i.e., agreed the most with) the statement that “PAIGE is useful in helping formulate a question about my health” (4.6). Patients gave the lowest mean score to (i.e., disagreed the most with) the statement that “Some of the questions from PAIGE used words that I didn’t understand” (1.4), which was a positive result for PAIGE.

**Exhibit 7.** Modified eHealth Literacy Scale Survey Results

Modified eHealth Question	Mean	Minimum	Maximum
PAIGE is <b>useful</b> in helping formulate a question about my health.	4.6	4	5
A system like PAIGE would save me time.	3.6	1	5
Some of the questions from PAIGE used words that I didn't understand.	1	1	1
PAIGE's questions were relevant follow-up questions in terms of my original question.	4.4	4	5
I feel that PAIGE provides high-quality information.	3.2	1	5
I thought it took PAIGE too long to generate each question.	1.4	1	3
Using a system like PAIGE would improve communication with my physician.	3.6	1	5

**4.1.3 Readability Results**

Exhibit 8 details the measures related to the reading level of PAIGE-generated text for the 10 patient interactions. PAIGE used an average of 292 words with a mean Flesch-Kincaid grade level of 7.2, indicating an average reading level of middle-school-aged children 11 to 14 years. This is in line with recommendations from the American Medical Association (AMA) for 6<sup>th</sup>- 8<sup>th</sup>-grade level and National Institutes of Health (NIH) for 8<sup>th</sup>-grade level.<sup>15</sup>

**Exhibit 8.** Flesch-Kincaid Grade Level for 10 Patient Interactions with PAIGE

Scenario	PAIGE-generated Text	
	Words	Flesch-Kincaid Grade Level
1	293	6.68
2	244	6.37
3	285	6.66
4	325	6.37
5	364	9.09
6	457	8.62
7	185	7.16
8	308	6.06
9	260	7.3

Scenario	PAIGE-generated Text	
	Words	Flesch-Kincaid Grade Level
10	216	7.2
<b>Mean</b>	<b>292</b>	<b>7.2</b>

#### 4.1.4 Key Qualitative Themes

Patients provided feedback on PAIGE regarding what they liked and thought could be improved. Generally, feedback was positive, but patients suggested both general and specific opportunities to make PAIGE more patient-centered and refine the prototype.

##### *What Patients Liked*

Most patients were enthusiastic about PAIGE, indicating it would be an effective tool for communicating with clinicians. Three themes emerged regarding what patients liked about PAIGE:

1. **PAIGE Produces Good Summaries.** All patients described the summaries generated by PAIGE positively, noting they were accurate, appropriate, and comprehensive. Informants were impressed by PAIGE’s ability to capture the essence of their questions and all details they provided about their symptoms. One patient liked that PAIGE “used the specificity” of what they typed in their original question. Some patients also liked the language and style of the summaries. For example, a patient liked that the tone of the summaries was not condescending, which is something they are on alert for when using patient-facing decision support.
2. **PAIGE Appropriately Uses AI.** Several patients reported that PAIGE appropriately used AI for clinical decision support. Specifically, they agreed with the decision to limit PAIGE to summarizing clinical questions rather than answering them. One patient input a hypothetical question about using a debunked alternative medicine therapy and noted that PAIGE “did the right thing” by simply conveying the question in the summary to the clinician, rather than giving advice for or against it. The patient noted the importance of not making anyone feel bad about asking questions.
3. **PAIGE Reduces Patient Burden.** Several patients noted that a major PAIGE benefit is the reduced burden on them to know what information to provide to a clinician. It may be especially helpful for patients with lower health literacy who could use support in understanding relevant symptoms or concerns to report. One patient noted that PAIGE could be used not only to support patients in reporting symptoms but also in reporting other things that may be impacting their health (e.g., air quality). Another patient described how it could be helpful for patients with conditions that cause brain fog who might not remember everything they need to tell their clinician.

## *Opportunities for Improvement*

Some patients reported aspects of PAIGE that could be improved. Specifically, patients noted four opportunities for improvement:

1. **PAIGE May Impact Personal Connection.** One patient noted that PAIGE felt like an “intermediary” that would reduce the personal connection with their clinician. The patient described that, due to their health conditions, they are highly engaged in their care and have developed strong relationships with their care teams. They believe calling their care team directly is the best strategy to receive the maximum, highest-quality care. The patient would feel like a barrier existed between themselves and their care team if they were required to communicate through a chatbot like PAIGE.
2. **Patient Communication Preferences Differ.** Similarly, one patient noted that a chatbot like PAIGE may not be the preferred communication method for all patients. In particular, the patient noted that patients who are highly engaged in their care may prefer more direct forms of communication like calling, while patients who are less engaged find PAIGE helpful.
3. **PAIGE May Not Save Patients Time.** While most patients thought PAIGE was an effective way to communicate their clinical concerns, they were mixed on whether it would be more efficient. Some were not convinced that it would be a quicker option compared to calling their clinician or messaging their clinician without the assistance of a chatbot. Others believed their clinician would ultimately end up asking them the same questions. However, patients did think that PAIGE could save clinicians’ time by reducing back-and-forth messaging to clarify patient questions and concerns. They noted this could ultimately result in a quicker response from their clinician.
4. **48-Hour Disclaimer is a Deterrent.** Most patients expressed dissatisfaction at the general disclaimer accompanying PAIGE-generated summaries, which notes that the clinician may take up to 48 hours to respond. They believed this timeframe was too long and would deter use of PAIGE, because patients would think calling would get them an answer faster.

## *Suggestions for Refinements*

Patients made three specific suggestions for refinements to PAIGE in future prototypes and two future efforts to undertake to improve PAIGE.

1. **Use the Patient’s Voice.** Though patients were generally satisfied with the PAIGE-generated summaries, some believed they could be improved by better using the patient’s voice/language. They noted that the summaries transformed the colloquial language the patient used in messages into formal language a clinician would use. For example, phrases like “no history of” or “I experienced a loss of voice” are more formal than how a patient would describe their symptoms. Patients noted they are more of an expert on their disease and experience than clinicians or AI, and worried that symptoms would get “lost in translation” if converted into more formal, clinical language.

2. **Provide Informational Resources.** Though patients supported PAIGE's focus on generating summaries, they also believed PAIGE could provide them with access to vetted clinical information or links to resources on the topics involved in their questions. Specifically, one patient suggested that PAIGE could provide optional resources to read while waiting for their clinician's response (e.g., evidence-based materials from HealthWise).<sup>16</sup> They noted that this would make PAIGE feel more like a service to patients, which could promote use.
3. **Incorporate Email Functionality.** One patient suggested adding functionality that would enable users to email the PAIGE-generated summary to themselves and/or to a family member or caregiver. They noted that this would enable family members or caregivers to review the summary for accuracy, which is particularly important in cases where conditions cause memory issues. It would also make PAIGE useful even if patients had a different communication preference (i.e., calling their clinician), as the patient could reference the email to efficiently describe their question and symptoms.

### *Suggestions for Future Work*

1. **More Patient Testing.** Patients suggested conducting future usability testing with patients with limited digital and/or health literacy. They noted that PAIGE may perform differently based on the level of detail patients use to describe their symptoms or the level of knowledge patients have about their symptoms or condition. Conducting future testing with patients with low digital and/or health literacy would support understanding of how PAIGE responds to differing styles and abilities to provide health information digitally. It could also inform understanding of how PAIGE responds when the patient speaks another language.
2. **EHR Data Integration.** Patients noticed that PAIGE asked questions about current medications and treatment history that could be avoided if PAIGE could reference patients' medical records. They noted that this capability—which is planned in future iterations—would assist PAIGE in asking more relevant questions and improve the overall experience of using PAIGE.

## 4.2 Assessment Goal 2: Quality and Utility of PAIGE-Generated Summaries

Exhibit 9 shows the summarized results of the clinician utility survey, in which clinicians were asked to rate their agreement with statements about the appropriateness, completeness, accuracy, and other aspects of 10 different interactions from the patient usability testing sessions. Clinicians could choose from a Likert scale where the strength of agreement increased from 1 to 5 (i.e., 1 meant strongly disagree with the statement and 5 meant strongly agree), and a score of 3 meant the clinician was neutral, or neither agreed nor disagreed. On average, clinicians agreed most with the statement about accuracy, i.e., that the PAIGE summaries accurately captured the patient's original questions and responses (4.7). On average, clinicians agreed least with the statement about completeness, i.e., that PAIGE's clarifying questions were complete and no other information would be required to respond to the patient's original question (3.2). The assessment also asked clinicians whether they preferred the patient's original question versus the PAIGE-generated summary or whether both were equally good. By far, clinicians chose the PAIGE summary (90% of clinicians on average across the 10 scenarios).

**Exhibit 9.** Clinician Ratings of PAIGE-Generated Questions and Summaries

Questions for Clinicians	Result
	<b>Mean Score</b>
PAIGE’s clarifying questions were <b>appropriate</b> given the patient’s original question.	4.5
PAIGE’s clarifying questions were <b>complete</b> (i.e., no other information would be required for me to respond) given the patient’s original question.	3.2
PAIGE’s clarifying questions requested <b>necessary</b> information to help me understand and respond to the patient’s original question.	4.2
The patient’s responses to the clarifying questions from PAIGE would be <b>useful</b> in helping me understand the patient’s problem and formulate a response.	4.3
The summary generated by PAIGE <b>accurately</b> captured the patient’s original question and responses to PAIGE’s clarifying questions.	4.7
Using a system like PAIGE would <b>reduce the time</b> required for me to answer patient questions.	4.3
	<b>Mean Percentage</b>
If given the choice between the patient’s original question and the summary question generated by PAIGE that included the patient’s answers to the PAIGE generated clarifying questions, which would you prefer?	
Patient’s Original Question	0.0%
PAIGE’s summary	90.0%
Both equally good	10.0%

Lastly, the assessment provided an open-ended text field where clinicians could provide additional context on their responses or additional thoughts and suggestions for improvement. Generally, clinicians reported PAIGE had potential with some improvements made to the questions and summaries, as well as to the broader health care delivery landscape.

#### 4.2.1 Suggestions to Improve PAIGE Queries

Aligning with the lower survey score on the completeness of PAIGE’s questions, clinicians made some specific recommendations for additional follow-up questions that PAIGE could ask to avoid the need for the clinician to seek more information. One clinician suggested that in a scenario related to a chronic autoimmune disease, PAIGE could ask about current treatment regimens and what treatment the patient has most recently received, noting that clinicians may not have this information in the EHR if the patient received treatment elsewhere. Another clinician noted that PAIGE in some cases should still ask

standard triage questions for common conditions (e.g., upper respiratory infection) even if the patient provided a medical history, as they are still the most likely causes. Similar to patient feedback, clinicians noted that PAIGE could ask more relevant follow-up questions if it had more patient context, such as age.

#### 4.2.2 Suggestions for PAIGE-Generated Summaries

To support clinicians in understanding whether PAIGE-generated summaries have all the information they need, one clinician specifically suggested that PAIGE include a one-liner with background information on the patient. For example, a patient's age is critical to conducting differential diagnosis. One clinician also noted that the summaries missed explicit pertinent negatives (e.g. no fever, no chills, no shortness of breath), even though the patient had explicitly denied having these symptoms. The clinician stated that, without knowing this information or that it was asked, they would have to ask the patient for the information again.

#### 4.2.3 Changes to Health Care Delivery

Multiple reviewers commented on clinicians' lack of time to diagnose and treat patients via portal messages, which often necessitates a reply of "I'm sorry to hear that. Please call my office so my staff can help you schedule a visit to discuss this further." One clinician noted that PAIGE holds promise once value-based care models and e-visits are more prevalent but may not be a time-saver now. The clinician noted that even as part of an accountable care organization, more than 90% of visits were paid fee for service. That said, one clinician reported that PAIGE has the potential in the meantime to decrease clinicians' cognitive load, which could be more impactful than the time burden.

#### 4.2.4 Discussion

Our findings indicate general enthusiasm for PAIGE from both patients and clinicians. Both end-user groups believed PAIGE held promise for enhancing patient-clinician communication for many common clinical concerns and questions. However, they also suggested improvements and future work to enhance PAIGE's information quality, better understand its benefit over current patient and clinician messaging workflows, and patient-centeredness.

**Patients find PAIGE easy to understand.** Our findings indicated positive results for our first main aim to assess patients' ability to comprehend and respond to PAIGE questions. No patients reported issues understanding PAIGE's questions and/or any terms used during usability testing sessions. Supporting this, patients disagreed most with the survey statement about PAIGE using words they didn't understand. Additionally, the Flesch-Kincaid grade level assessment revealed that PAIGE aligned with recommendations from the AMA and NIH (i.e., used words with a reading level appropriate for middle-school-aged children 11 to 14 years).

**Clinicians find PAIGE useful.** Results were also positive for our second main aim to assess the utility and quality of PAIGE-generated summaries for clinicians. In general, clinicians agreed that PAIGE



asked appropriate and necessary clarifying questions and generated accurate summaries with useful information. By far, clinicians opted for PAIGE's summary over the patient's original question.

**PAIGE may not save users' time.** Despite these results, some patients and clinicians expressed uncertainty that PAIGE would save them time. For patients, this stemmed from not feeling confident that PAIGE would be quicker than calling their clinician, especially because of the 48-hour response window. For clinicians, this stemmed from having limited time for portal messages of any kind (even those summarized by a chatbot) under the existing care delivery structure. Future testing should examine whether PAIGE has time-saving capabilities for both end-user groups.

**PAIGE would benefit from contextual information about patients.** Both groups also remarked on the importance of PAIGE having more contextual information about the patient, such as their demographics, medications, and treatment history. This underscores the importance of building EHR integration into future prototypes and conducting further testing to assess how PAIGE handles this information and whether clinicians' opinions about PAIGE's time-saving capabilities change.

**Refinements to increase PAIGE patient-centeredness.** Finally, patients provided important feedback to ensure patient-centeredness in future iterations of the prototype, including adding email functionality, providing informational resources, and using more of the "patient voice" in summaries for clinicians. They also recommended conducting testing with patients with limited digital or health literacy to understand whether PAIGE is effective for different patient populations.

The generally positive reception of PAIGE among patients and clinicians is an optimistic sign for developers of similar tools that use AI to interact with patients. However, the suggestions for improvements and future work reveal important considerations for the design and success of such tools. In particular, the PAIGE assessment revealed that some uncertainty exists among patients and clinicians about the benefit of using AI-based chatbots to support their interactions. Our assessment begins to provide insight into the specific features of AI-based chatbots that will enable them to save time and reduce cognitive burden for both clinicians and patients, as well as feel like a service to patients rather than an extra step in the communication process. For example, patient recommendations to improve patient-centeredness indicate that providing educational information and integrating functionalities (e.g., email) that empower use of health information are important ways AI-based chatbots can offer more utility to patients. However, more research is necessary to understand whether and how AI-based chatbots can streamline communication between clinicians and patients, and which clinician types and patient populations will benefit most from such tools.

## 5. Conclusion

Conversational AI chatbots have the potential to support and streamline communication between patients and clinicians, but such technologies must be designed carefully with these users in mind. As an early prototype of such a chatbot, PAIGE shows promise at being an appropriate, usable, and useful AI-based tool for both end-user groups. Developers will continue to refine and test PAIGE based on the feedback gathered in this assessment to ensure future prototypes enhance usability and utility for different patient and clinician populations. Others creating similar tools can apply the feedback on PAIGE to enhance and refine their prototypes.

## Appendix A: Example Interaction Between a Patient and PAIGE

**Patient's Original Question:** I started taking Arikayce last week and I'm losing my voice. What should I do?

**PAIGE:** This is PAIGE. I'm an AI helper that will assist you in sending a message to your physician. I will ask you some follow-up questions to clarify important information and help ensure your physician will be able to accurately answer your question the first time.

To start, could you tell me how long after starting Arikayce did you begin to notice changes in your voice?

Patient: 2 days

**PAIGE:** Thank you for that detail. Have you experienced any other symptoms besides the change in your voice, such as sore throat, cough, or difficulty breathing?

**Patient:** my throat is a little scratchy

**PAIGE:** Got it, a scratchy throat along with the voice change. Have you had any fever or felt unwell in other ways since these symptoms started?

**Patient:** no not really

**PAIGE:** That's good to hear. To help your physician understand your current medication routine, can you confirm if you are taking any other medications or have any other health conditions that your physician should be aware of?

**Patient:** yes, I'm taking azithromycin, ethambutol and a third antibiotic for my MAC. I have a MAC infection and bronchiectasis.

**PAIGE:** Thank you for sharing your medication list and health conditions. One last question: How has your voice change affected your daily activities or ability to communicate?

**Patient:** it is harder to speak to people because I am losing my voice

**PAIGE:** Thank you for providing all this information. Here's the summary of your message that we'll send to your physician:

Dear Doctor,

I started taking Arikayce last week for my MAC infection, and two days after starting the medication, I began experiencing a loss of voice and a scratchy throat. I have not had any fever or other significant new symptoms. Currently, I am also on azithromycin and ethambutol for MAC, and I have a history of bronchiectasis. The change in my voice has made it challenging for me to communicate effectively in my daily activities. Could you please advise on how I should proceed?

## Appendix B. Modified eHealth Literacy Scale

I would like to ask you for your opinion on using PAIGE for communicating with your physician. For each statement, tell me which response best reflects your opinion and experience.

1. PAIGE is useful in helping formulate a question about my health.

1	2	3	4	5
Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree

2. A system like PAIGE would save me time.

1	2	3	4	5
Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree

3. Some of the questions from PAIGE used words that I didn't understand.

1	2	3	4	5
Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree

4. PAIGE's questions were relevant follow-up questions in terms of my original question.

1	2	3	4	5
Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree

5. I feel that PAIGE provides high-quality information.

1	2	3	4	5
Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree

6. I thought it took PAIGE too long to generate each question.

1	2	3	4	5
Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree

7. Using a system like PAIGE would improve communication with my physician.

1	2	3	4	5
Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree

## Appendix C. Instrument for Clinician Assessment

We are developing a chatbot system called Patient Artificial Intelligence Guided E-Messages (PAIGE) that is aimed at improving patient-clinician communication. Leveraging a large language model, PAIGE is intended to serve as an intermediary between patients and clinicians within the patient portal to (1) ask clarifying questions about symptoms on the clinician's behalf and (2) send a summary of the patient's question and their responses to the clinician to facilitate decision-making. PAIGE is specifically designed NOT to provide medical advice. At the current phase of development, we are interested in getting the opinion of clinicians on the utility of such a tool. For this short study, we are asking you to focus on PAIGE's ability to generate clear, understandable, and relevant questions for patients. We are not focusing on the usability or the "look and feel" of the user interface, as it is not yet integrated into the patient portal.

Following each of the 10 example patient exchanges with PAIGE, we have 7 short questions for you to answer about the exchange. We expect this entire assessment to take less than 30 minutes.

[example patient exchange here]

**For each statement, choose the response that best aligns with your view for the exchange above.**

1. PAIGE's clarifying questions were appropriate given the patient's original question.

1	2	3	4	5
Strongly Disagree	Disagree	Neither Agree Nor Disagree	Agree	Strongly Agree

2. PAIGE's clarifying questions were complete (i.e., no other information would be required for me to respond) given the patient's original question.

1	2	3	4	5
Strongly Disagree	Disagree	Neither Agree Nor Disagree	Agree	Strongly Agree

3. PAIGE's clarifying questions requested necessary information to help me understand and respond to the patient's original question.

1	2	3	4	5
Strongly Disagree	Disagree	Neither Agree Nor Disagree	Agree	Strongly Agree

4. The patient's responses to the clarifying questions from PAIGE would be useful in helping me understand the patient's problem and formulate a response.

1	2	3	4	5
Strongly Disagree	Disagree	Neither Agree Nor Disagree	Agree	Strongly Agree

5. The summary generated by PAIGE accurately captured the patient's original question and responses to PAIGE's clarifying questions.

1	2	3	4	5
Strongly Disagree	Disagree	Neither Agree Nor Disagree	Agree	Strongly Agree

6. Using a system like PAIGE would reduce the time required for me to answer patient questions.

1	2	3	4	5
Strongly Disagree	Disagree	Neither Agree Nor Disagree	Agree	Strongly Agree

7. If given the choice between the patient's original question and the summary question generated by PAIGE that included the patient's answers to the PAIGE generated clarifying questions, which would you prefer?

Patient's Original Question	PAIGE's Summary	Both Equally Good
-----------------------------	-----------------	-------------------

## References

- <sup>1</sup> Chavez A, Bracamonte J, Kresin M, Yardley M, Grover M. High Volume Portal Usage Impacts Practice Resources. *J Am Board Fam Med* 2020;33:452–5. <https://doi.org/10.3122/jabfm.2020.03.190401>.
- <sup>2</sup> Reed M, Graetz I, Gordon N, Fung V. Patient-initiated e-mails to providers: associations with out-of-pocket visit costs, and impact on care-seeking and health. *Am J Manag Care* 2015;21:e632-639.
- <sup>3</sup> Adler-Milstein J, Zhao W, Willard-Grace R, Knox M, Grumbach K. Electronic health records and burnout: Time spent on the electronic health record after hours and message volume associated with exhaustion but not with cynicism among primary care clinicians. *J Am Med Inform Assoc* 2020;27:531–8. <https://doi.org/10.1093/jamia/ocz220>.
- <sup>4</sup> Arndt BG, Beasley JW, Watkinson MD, Temte JL, Tuan W-J, Sinsky CA, Gilchrist VJ. Tethered to the EHR: Primary Care Physician Workload Assessment Using EHR Event Log Data and Time-Motion Observations. *Ann Fam Med* 2017;15:419–26. <https://doi.org/10.1370/afm.2121>.
- <sup>5</sup> Akbar F, Mark G, Warton EM, Reed ME, Prausnitz S, East JA, Moeller MF, Lieu TA. Physicians' electronic inbox work patterns and factors associated with high inbox work duration. *J Am Med Inform Assoc* 2021;28:923–30. <https://doi.org/10.1093/jamia/ocaa229>.
- <sup>6</sup> CGS Survey Reveals Consumers Prefer a Hybrid AI/Human Approach to Customer Service. Is there Chatbot Fatigue?. CGS. 2019. URL: <https://www.cgsinc.com/en/resources/2019-cgs-customer-service-chatbots-channels-survey> (Accessed 12 December 2023).
- <sup>7</sup> Liu J, Wang C, Liu S. Utility of ChatGPT in Clinical Practice. *J Med Internet Res* 2023;25:e48568. <https://doi.org/10.2196/48568>.
- <sup>8</sup> Milne-Ives M, de Cock C, Lim E, Shehadeh MH, de Pennington N, Mole G, Normando E, Meinert E. The Effectiveness of Artificial Intelligence Conversational Agents in Health Care: Systematic Review. *J Med Internet Res* 2020;22:e20346. <https://doi.org/10.2196/20346>.
- <sup>9</sup> Tudor Car L, Dhinakaran DA, Kyaw BM, Kowatsch T, Joty S, Theng Y-L, Atun R. Conversational Agents in Health Care: Scoping Review and Conceptual Analysis. *J Med Internet Res* 2020;22:e17158. <https://doi.org/10.2196/17158>.
- <sup>10</sup> Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, Faix DJ, Goodman AM, Longhurst CA, Hogarth M, Smith DM. Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum. *JAMA Intern Med* 2023;183:589–96. <https://doi.org/10.1001/jamainternmed.2023.1838>.
- <sup>11</sup> Liu S, McCoy AB, Wright AP, Carew B, Genkins JZ, Huang SS, Peterson JF, Steitz B, Wright A. Leveraging Large Language Models for Generating Responses to Patient Messages 2023:2023.07.14.23292669. <https://doi.org/10.1101/2023.07.14.23292669>.

- <sup>12</sup> Clinical Decision Support Innovation Collaborative. Accessed July 2, 2024. <https://cdsic.ahrq.gov/>
- <sup>13</sup> Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. *J Med Internet Res*. 2006 Nov 14;8(4):e27. doi: 10.2196/jmir.8.4.e27. PMID: 17213046; PMCID: PMC1794004.
- <sup>14</sup> Flesch R. A new readability yardstick. *J Appl Psychol*. 1948;32(3):221-233. doi:10.1037/h0057532.
- <sup>15</sup> Rooney MK, Santiago G, Perni S, et al. Readability of Patient Education Materials From High-Impact Medical Journals: A 20-Year Analysis. *J Patient Exp*. 2021;8:2374373521998847. Published 2021 Mar 3. doi:10.1177/2374373521998847
- <sup>16</sup> Healthwise by WebMD Ignite. Accessed July 23, 2024. <https://www.healthwise.org/>