This study aims to explore healthcare provider acceptability of AI documentation. Generative AI, like OpenAI’s GPT, has led to a surge in its adoption, especially during the COVID-19 pandemic, to alleviate documentation-related burnout by automating tasks such as generating replies to patient inquiries, drafting clinic letters, and assisting in medical notes. This study examines the capability of AI-generated patient portal messaging, offering insights into the feasibility of integrating large language models (LLMs) to enhance efficiency in patient communication and alleviate workload burdens on clinicians.

**Methodology**

**Initial Synthetic Patient-Message Data Collection**
1. Retrieved 85 patient portal messages and clinician responses from Vanderbilt University Medical Center (VUMC) repository.
2. De-identified and manually rephrased messages to maintain content while varying tone and length.
3. Engineered a prompt within GPT-4 to generate messages similar in tone, length, and topic to the originals.
4. Recruited eight clinicians to review and compare synthetic and authentic patient portal messages, finding that clinicians correctly identified clinician-generated messages only 51.1% of the time.
5. Combined the message pool to develop synthetic patient portal message responses based on the study’s findings.

**Pipeline Development**

- **Initial Prompt**
  - PROPOSED MESSAGE

- **Change Ganmen**
  - As a generative checking AI, your role is to identify and correct any grammatical errors, missing information, or incoherent texts. Please produce a corrected version of the text.

- **Literacy Level**
  - As an AI assistant, it’s important to consider the literacy level of the patient. The categories are middle school, high school, or higher education.

- **Category**
  - As an AI assistant, you will be responsible for generating the appropriate category based on the content. You have the following categories: (1) General, (2) Medical, (3) Urgency, (4) Relevance, (5) Empathy, (6) Readability, (7) Accuracy, (8) Relevance, (9) Medical Accuracy

- **Generate Response**
  - You are an urgent classifier. Classify the urgency of the following message and assign it to one of the following categories: (1) High, (2) Medium, (3) Low.

- **Urgency Check**
  - Confusion of urgency; identify the urgency level of the message. Your job is to identify the correct urgency level and assign a corresponding category.

- **App Library**
  - As an AI assistant, you will be responsible for generating responses to patient messages. Your job is to draft the appropriate response and assign it to one of the following categories: (1) General, (2) Medical, (3) Urgency, (4) Relevance, (5) Empathy, (6) Readability, (7) Accuracy, (8) Relevance, (9) Medical Accuracy.

**Evaluation of Message Response Pairs**

- **Patient**
  - The pain gets worse when I’m in a bright room or when I try to do basket inquiries, drafting clinic letters and assisting in medical notes.

- **Doctor**
  - Let’s schedule an appointment to discuss. In the meantime, keep a diary of what you are eating and drinking, and we can review that.

- **Patient**
  - Sounds like your headaches are very severe. Let’s schedule an appointment to discuss. In the meantime, keep a diary of what you are eating and drinking, and we can review that.

**Results**

- **Response Assessment: Real vs. GPT**
  - Figure 2: Evaluation of the Pipeline. The radar diagram illustrates the mean comparison of GPT-generated and real responses using an ordinal scale ranging from (1) low to (5) high. A rating of 1 indicates poor performance, while 5 signifies excellent performance.

**Table 1:** Comparative Analysis of GPT versus real message responses. The table above provides a comprehensive breakdown of the average means and medians derived for the four key characteristics, comparing GPT-generated message-response pairs to real ones. Both empathy and readability were statistically better for GPT-generated responses.

**Conclusions**

The findings of this study suggest that GPT-4 generated responses are feasible and acceptable to primary care providers. Despite the small sample size and single healthcare system representation, the study provides promising insights into the potential of AI-driven messaging systems to alleviate clinician burnout and enhance patient communication. As with all technological endeavors, continual evolution is paramount for addressing challenges and leveraging emerging insights from both the technological and cognitive domains.

**References**