scientific reports

Check for updates

OPEN

Enhancing patient participation in emergency department through patient-friendly clinical notes generated by large language models

Sung-In Kim[1], Joonyoung Park[2], Taewan Kim[3], Woosuk Seo[4], Taerim Kim[5], Won Chul Cha[5] & Hwajung Hong[2,6]✉

Patient-centered care (PCC) emphasizes providing patients with clear information to support active participation in medical decision-making. However, the fast-paced nature of emergency departments (ED), coupled with communication barriers and varying health literacy, limits effective patient engagement. While large language models (LLMs) have shown potential in generating patient-friendly documents, their use in ED settings remains underexplored. This study aimed to develop LLM-generated patient-friendly clinical notes (PFCNs) that transform clinical notes into plain language, and to evaluate whether PFCNs could enhance patient participation in ED consultations. In this study, a total of 120 PFCNs were generated and evaluated, receiving high understandability ratings from both 10 clinicians and 20 patients (PEMAT score: 87.2%). Patients who used PFCNs during simulated ED consultations reported significantly higher participation levels compared to prior ED experiences (PPQ, $P < 0.05$). Qualitative data showed that PFCNs supported understanding, question preparation, emotional reassurance and improved relationships with clinicians, though concerns about hallucinations and integration into clinical workflows remained. These findings suggested that PFCNs generated by LLMs show promise for enhancing patient participation in ED consultations. Future work should address accuracy and explore real-world integration to support safe and effective deployment.

Keywords Patient-centered care, Patient participation, Large language model, Emergency medicine, Communication

Patient-centered care (PCC) emphasizes active participation of patients in the medical decision-making process by aligning care with their values and preferences[1]. Central to PCC is the promotion of patient autonomy, which requires active patient participation in various aspects of care, such as shared decision-making and communication with healthcare providers[2–4]. While patient participation has been linked to improved understanding[5], emotional well-being[6,7], and clinical outcomes[8], it remains difficult to achieve in emergency department (ED) settings. The fast-paced environment, limited time, and differences in communication styles between clinicians and patients often result in fragmented communication and reduced patient involvement[1,9,10].

To promote participation of patients in clinical settings, several strategies have been developed to provide patients with accessible medical information[11], including chatbot-based informed consent[12] and patient-friendly discharge documents[13]. Recently, large language models (LLM) have emerged as promising tools to generate plain-language summaries of clinical information[13–15]. By translating clinical notes into more understandable formats, LLMs have the potential to bridge information gaps and support patient participation.

[1]Department of Psychiatry, Seoul National University College of Medicine, Seoul, Republic of Korea. [2]Department of Industrial Design, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea. [3]TwelveLabs, Seoul, Republic of Korea. [4]Department of Emergency Medicine, Yale School of Medicine, Connecticut, USA. [5]Department of Emergency Medicine, Samsung Medical Center, Seoul, Republic of Korea. [6]Korea Advanced Institute of Science and Technology (KAIST), 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea. ✉email: hwajung@kaist.ac.kr

However, the application of LLM in high-pressure environments like the ED raises important concerns because generative AI systems are prone to "hallucinations," or producing inaccurate or misleading content[13,16,17]. Recent studies on LLM use in clinician–patient communication have shown that such hallucinations can manifest as omitted key information, fabricated clinical details, or misinterpretation of medical terms[18,19]. In the context of clinical conversations, hallucinated or incorrect information can erode patient trust, disrupt shared decision-making, and potentially lead to harmful medical decisions[20]. Particularly in ED consultations—where medical decisions are often time-sensitive and emotionally charged—such inaccuracies could lead to patient confusion, decreased trust, or even safety and legal risks[21,22]. Given real-world deployment of LLM-generated documents requires careful evaluation, in-lab simulation studies offer a safe environment to assess both the potential and limitations of LLMs in ED consultations before implementation in practice[23,24].

This study aims to assess whether LLM-generated patient-friendly clinical notes (PFCNs) can support patient participation in ED consultations. We first conducted interviews with patients and clinicians to identify appropriate content, format, and target patients for PFCNs in ED settings. Based on these findings, we developed a system using GPT-4o to transform clinicians' notes into patient-friendly documents. We then conducted a simulated ED consultation role-play study involving patients and clinicians to identify whether PFCNs could promote participation, and challenges such as hallucinations might arise in the ED context. By identifying both opportunities and risks of PFCNs in ED settings, this study seeks to inform the future implementation and integration of generative AI tools into ED workflows.

Methods

Ethical considerations & recruitments

This study protocol and participant recruitment plan were approved by the Institutional Review Board (IRB) of Samsung Medical Center (IRB No. 2024-11-023-001). The inclusion criteria for clinician participants were: (1) adults aged 18 or older, (2) those with experience in ED care, (3) individuals capable of effective communication, and (4) those who agreed to the purpose and content of the study. The inclusion criteria for patient participants were: (1) adults aged 18 or older, (2) individuals with personal experience receiving ED care, (3) those capable of effective communication, and (4) those who agreed to the purpose and content of the study.

We aimed to recruit 20 physician participants and 30 patient participants. Recruitment was conducted by posting announcements on the offline bulletin boards of Samsung Medical Center and on an online community platform Everytime[25] dedicated to Seoul National University students. Recruited participants were informed of the study's purpose and methods and were asked to provide written informed consent. Each participant who completed the study received a 20 USD compensation. All methods were performed in accordance with the relevant guidelines and regulations.

Preliminary study: interview for developing PFCNs using LLM

We conducted individual interviews with nine patients and seven clinicians to explore communication challenges and user expectations for PFCNs in ED settings within one hour.

Through the interview results, we identified the need for PFCNs to bridge the information gap between clinicians and patients in the ED setting. Many patients reported difficulty understanding medical information due to the fast-paced environment and the complex medical terminology often used by clinicians. On the other hand, clinicians found it challenging to provide detailed explanations to fulfill patients' information needs due to their heavy workload. Given these challenges, both patients and clinicians expressed support for incorporating PFCNs that include the patient's condition, treatment process, and clinical information (e.g. diagnosis, test results) into the ED consultation process.

Participants highlighted that medical information was often conveyed verbally, in complex terms, and without adequate explanation—creating barriers to patient understanding. Therefore, both patients and clinicians recommended that PFCNs be written in plain language, with key medical terms simplified or supplemented with additional explanations. They also emphasized the importance of clear headings (e.g., Diagnosis, Treatment Plan) and visual cues (e.g., bold or underline) to help patients easily locate important information. Clinicians reported three key moments in the ED workflow where PFCNs could be particularly useful: after the initial consultation, during follow-up, and at discharge—particularly to address patient uncertainty during long waiting times.

However, concerns were raised regarding the feasibility of using PFCNs with patients in severe or unstable conditions (e.g., stroke, acute myocardial infarction), or those with low health literacy. To address this, we focused our subsequent evaluation on patients with milder conditions, who were more likely to engage with and benefit from the PFCNs.

Development of the LLM-based PFCNs

Based on the results of the preliminary study, we developed prompts to generate PFCNs using clinicians' clinical notes after patient consultations. We selected GPT-4o[26] as the LLM, as it was the most up-to-date model available in July 2024 and capable of supporting the Korean language. Prompts were tailored for three stages: initial visit, follow-up, and discharge. We incorporated three key considerations identified from the preliminary study interview findings into the prompts:

1. Improving accuracy.

- Interview findings: Patients consistently emphasized that PFCNs should closely match their clinicians' verbal explanations, noting that discrepancies could cause confusion and reduce trust. Clinicians, particularly those with prior LLM experience, highlighted the risk of hallucinations—where fabricated or inaccurate information could lead to misunderstandings or even legal issues. Both groups stressed the importance of basing PFCN content strictly on clinical notes and verbal consultations.

- Prompt design: To address these concerns, we developed accuracy-focused prompts that explicitly instructed the model to avoid introducing any content not documented in the clinical notes. Context-based instructions (e.g., "This document is written by a clinician and provided to patients in the emergency department.") clarified the clinical setting, while error-handling constraints (e.g., "Avoid including any information not explicitly documented in the clinical notes") were integrated to minimize hallucinations.

2. Enhancing patient-friendliness.

- Interview findings: Both patients and clinicians emphasized the need for PFCNs to be understandable to individuals with low health literacy. Participants recommended replacing medical jargon with plain language, providing supplemental explanations for complex terms, and including mechanisms for patients to ask follow-up questions. Clinicians noted that such features could help clarify misunderstandings and foster two-way communication.
- Prompt design: We incorporated instructions to convert technical medical terms into plain language and to append brief, accessible explanations when necessary. The prompts also specified inclusion of a greeting and instructions for the "Question" function, enabling patients to submit questions directly to their clinicians for clarification.

3. Improving readability.

- Interview findings: Participants suggested that clear visual structure would help patients locate and recall important information. Recommendations included organizing the content with subheadings, highlighting critical information (e.g., diagnoses, test names), and using formatting such as bold or underlined text for emphasis.
- Prompt design: The prompts instructed the LLM to format PFCNs with clear section headings (e.g., "Diagnosis", "Future Treatment Plan") and to use bold text to highlight key terms. These formatting guidelines were designed to improve scanability and ensure that patients could quickly identify the most relevant details.

Prompts used zero-shot learning to accommodate varied documentation styles and included specific rules for each document type—for instance, discharge PFCNs emphasized post-care instructions. The full prompt structure is available in Supplement File 1.

To facilitate the generation and delivery of PFCNs during the ED consultations process, we designed a LLM-based system embedded with the prompts. This system enables clinicians to input clinical notes, generate PFCNs, and send them to patients. Patients can be provided the documents and submit questions which clinicians can address during subsequent consultations (Fig. 1). As of July 2024, the system was developed using Next.js (version 15)[27] with TypeScript, HTML, and CSS. On the clinician-side interface, clinicians can select a medical note and its timestamp, generate PFCNs, and send finalized documents to patients (Fig. 2). On the patient-side interface, patients can view the PFCNs sent by their clinicians and submit questions, which are displayed on the clinician interface (Fig. 3).

Evaluation of PFCNs

Phase 1: clinician evaluation

We conducted an evaluation with 10 clinicians to assess the quality and applicability of PFCNs in ED (Fig. 4). Each clinician participated in two role-play scenarios (Table 1) with a simulated patient (SIK), covering three consultation stages: initial visit, follow-up, and discharge. After each consultation, clinicians documented the

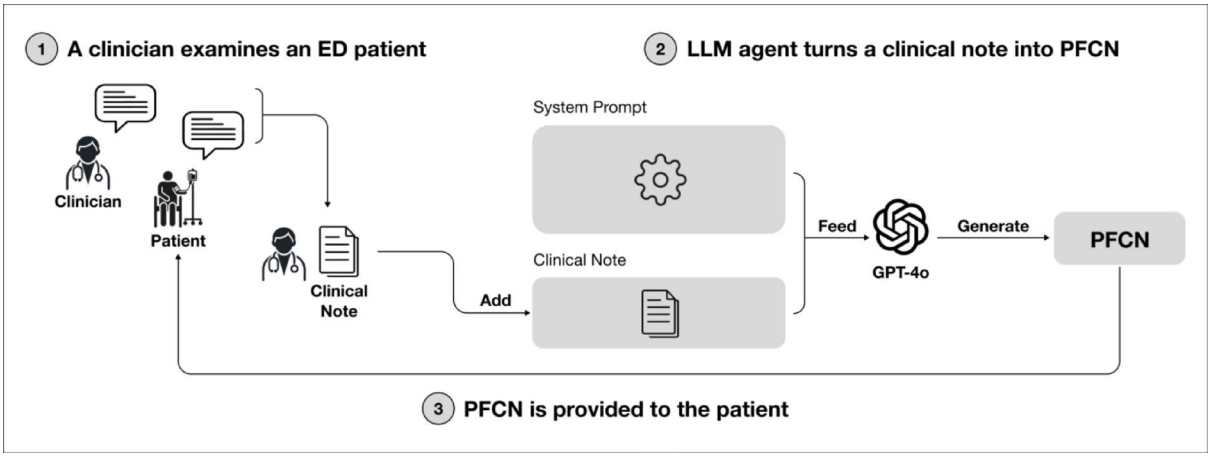

Fig. 1. The process of generating and utilizing PFCNs: (1) The clinician writes clinical notes after the patient consultation. After that, the clinician adds these clinical notes to the LLM-based system. (2) The system processes the clinical notes into PFCNs, along with the system prompts developed by the researchers. (3) The PFCNs are delivered to the patient and continue to be used in subsequent ED experiences.

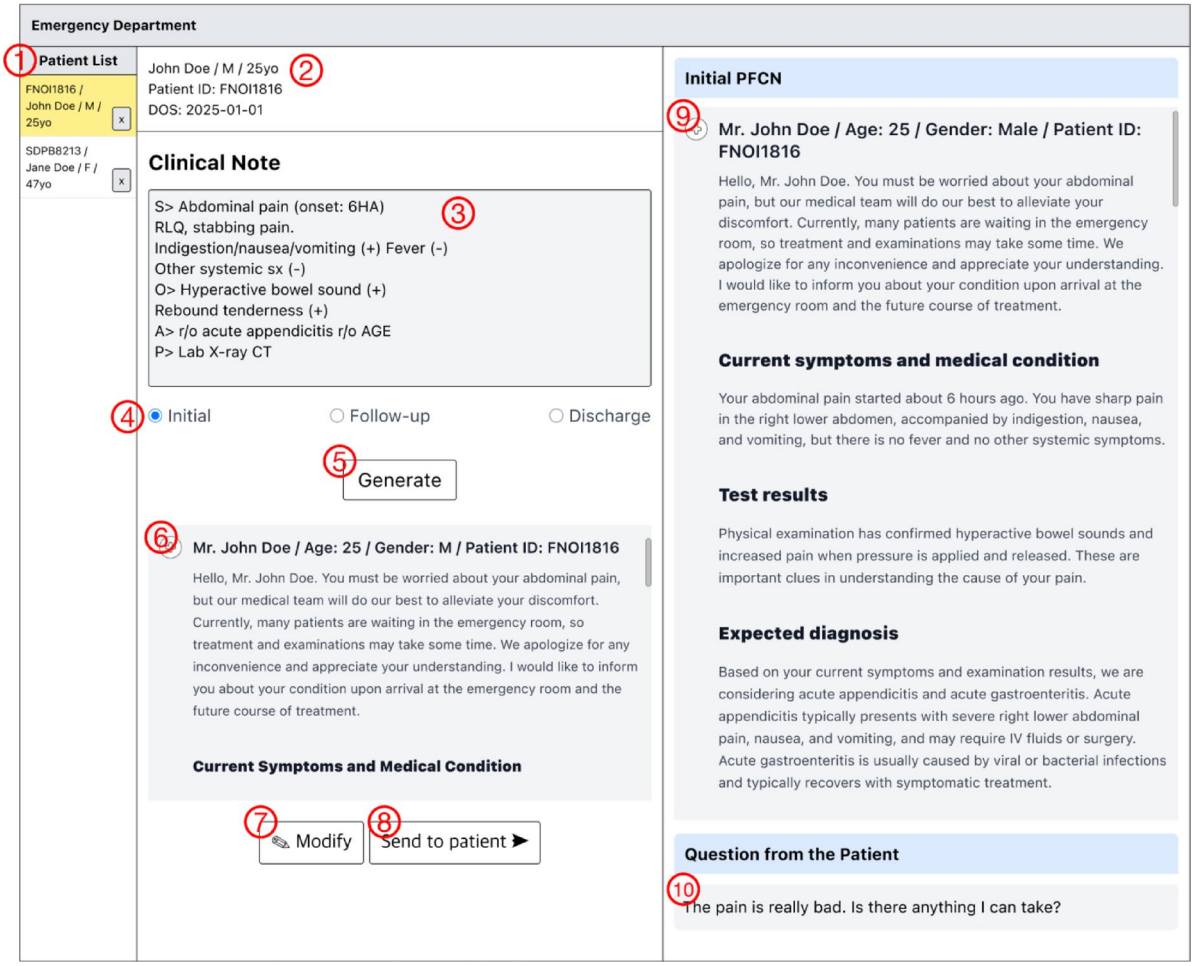

Fig. 2. Clinician interface for generating PFCNs: (1) List of current patients in the ED. (2) Personal information of the selected patient from the list. (3) Text area for entering clinician's clinical note. (4) Buttons for selecting one of three consultation processes, which determines the type of PFCN to be generated. (5) Button to generate PFCNs from the clinical note. (6) Panel for displaying generated PFCNs. (7) Button to modify generated PFCNs. (8) Clicking the "Send to patient" button delivers the PFCNs to the patients' interface. (9) Patients can view the same content in Fig. 3-(2). (10) Panel for displaying questions received from the patients.

encounter in clinical notes, which were then used to generate PFCNs. To observe changes in clinician-patient communication following document delivery, the simulated patient asked questions about the document's contents during the consultation.

Clinicians evaluated six PFCNs based on three criteria: perceived accuracy, completeness, and understandability [Supplement File 2]. Perceived Accuracy was assessed on a 6-point Likert scale to assess how accurately the document reflected the contents of the clinical note[13]. Participants also explained the reasons for their accuracy ratings and commented on the document content. Completeness was assessed by checking if the generated PFCNs reflected all essential elements from the clinical notes. Participants were asked to note any missing information in the completeness evaluation form. Lastly, Understandability was measured using the Patient Education Materials Assessment Tool (PEMAT), focusing on the 16-item understandability Sect[28]. [Supplement File 2]. A score of 70% or higher is generally considered indicative of adequate understandability.

Following the evaluation, individual interviews were conducted with clinicians to gather their feedback on the PFCNs review process, communication during consultation, suggestions for improvement, and potential implementation strategies. Participants were informed that the interviews would be recorded, anonymized, and analyzed. They were encouraged to freely share their thoughts on the use of the PFCNs and were assured that they could withdraw from the interview at any time.

Phase 2: patient evaluation

We conducted an evaluation with 20 patients to assess their perspectives on PFCNs (Fig. 4). Each participant engaged in three simulated consultations (initial, follow-up, discharge) with a clinician (SIK), presenting symptoms based on their recent ED visits. After each consultation, they received PFCNs on a tablet and were encouraged to review the documents and ask clarifying questions during subsequent sessions. Participants

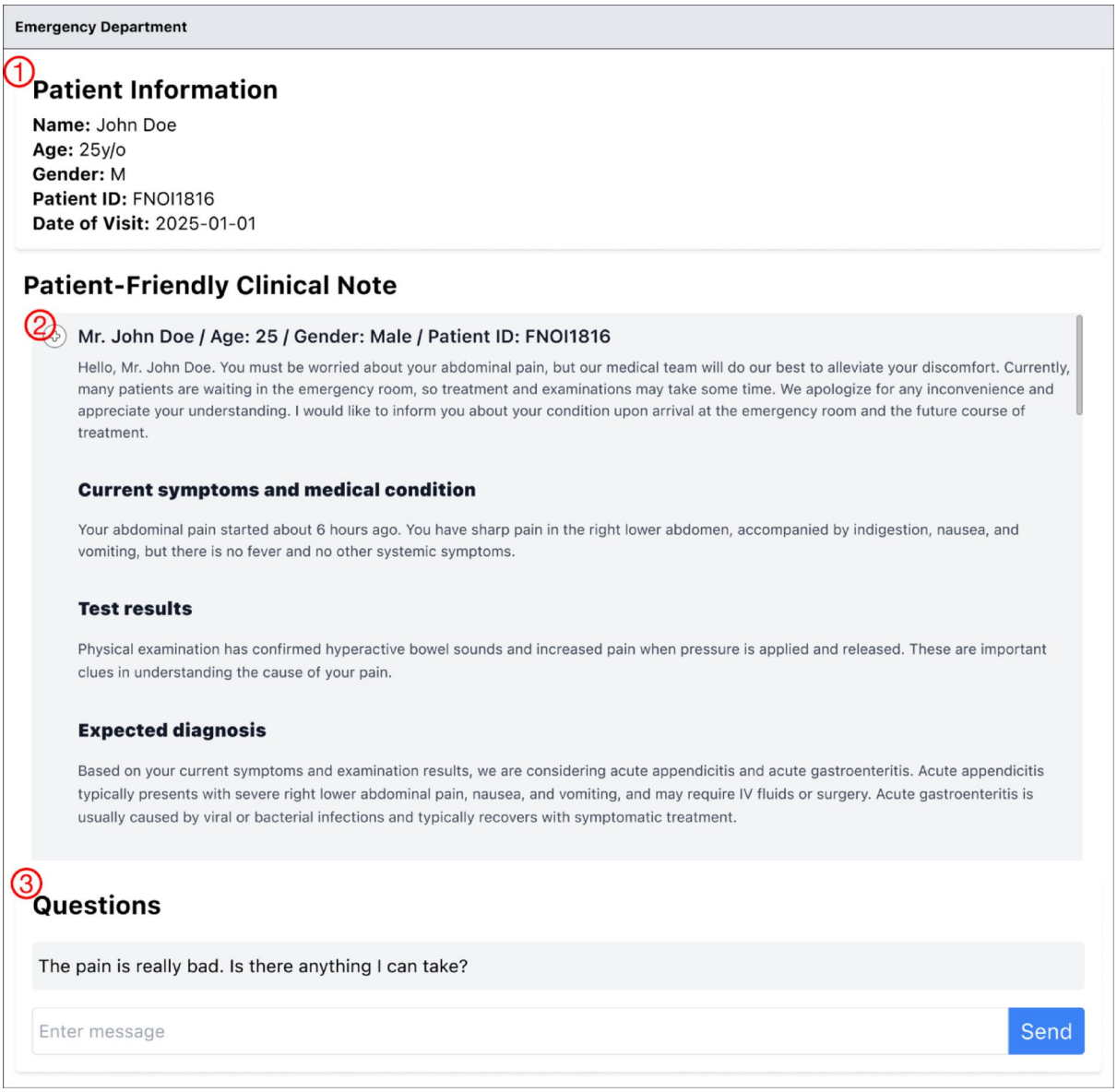

Fig. 3. Patient interface for reading PFCN and sending questions: (1) Panel displaying the patient's personal information. (2) Panel showing the PFCN sent by the clinician. (3) Panel for sending questions to the clinician. After writing questions, the patient clicks the "Send" button, and the questions appear in the clinician's interface (Fig. 2-(10)).

provided informed consent for session recordings and were assured they could withdraw at any time if they experienced discomfort or concerns during the study.

Before and after the simulated consultations, patient participation was assessed using the Patient Participation Questionnaire (PPQ)[2], a 17-item tool measuring perceived involvement in decision-making and communication [Supplement File 2]. In the study, patients completed the PPQ to evaluate their participation in previous ED experiences before the simulation and then evaluated their participation during the simulation using the PFCNs.

After the simulation, to evaluate the usefulness of the PFCNs, patients completed the 9-item Usefulness Scale for Patient Information Material (USE)[29], which assesses cognitive, emotional, and behavioral support provided by the documents [Supplement File 2]. Understandability was also assessed using PEMAT understandability Sect[28].

To further explore patients' experience using PFCNs, 30-minute interviews were conducted to understand how the documents supported their understanding and participation, as well as ideas for improving and implementing such tools in real ED settings. Participants were informed that the interviews would be recorded, anonymized, and analyzed, and were encouraged to speak freely with the assurance that they could withdraw at any time.

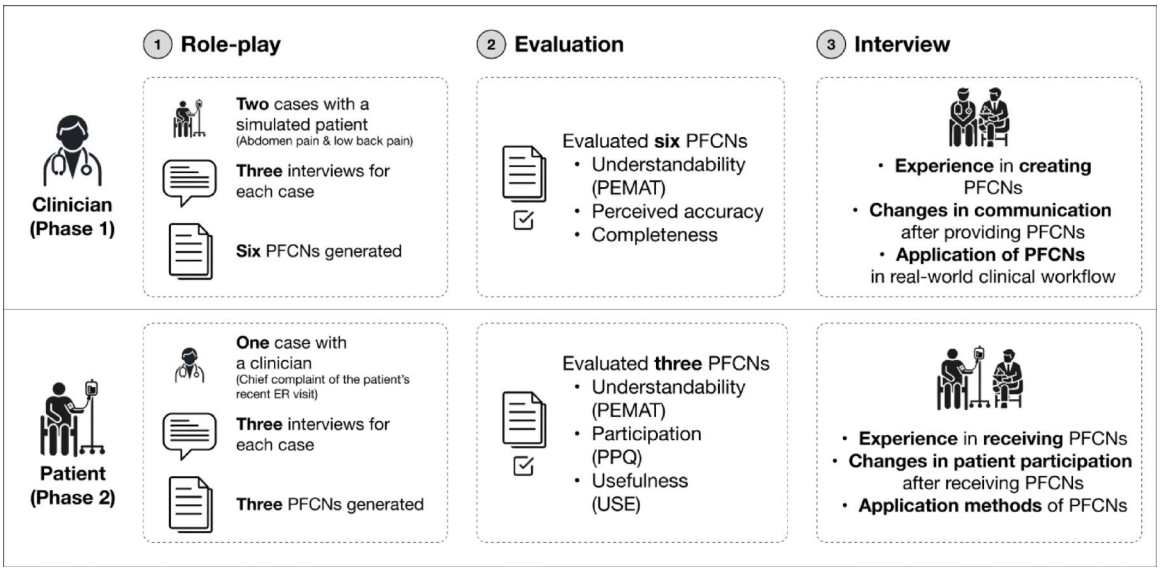

Fig. 4. Evaluation process of LLM-generated PFCNs of clinicians and patients.

Scenario	Chief complaint	Descriptions
#1	Abdomen pain	Chief complaint : Right Lower abdomen pain Onset : 6 h ago Subjective symptoms : Right Lower abdomen pain, abdomen discomfort, dyspepsia Physical findings : Tenderness(+), Rebound tenderness (-) Previous medical history : n/a
#2	Low back pain	Chief complaint : low back pain Onset : 3 h ago Subjective symptoms : Low back pain, Radiating leg pain, numbness Physical findings : straight leg raise test. right (+) Previous medical history : herniated intervertebral disk in L5-6, L6-7

Table 1. Simulated role-play scenarios for clinicians in ED consultations.

Analysis

For quantitative analysis, the PPQ data collected during the study were analyzed using the open-source data analysis tool Jamovi[30]. Changes in PPQ scores before and after the role-play sessions were examined. Since the normality test indicated a non-normal distribution ($p < 0.05$), we used the Wilcoxon Signed-Rank Test to compare pre- and post-study PPQ scores. The results of the PEMAT and USE scales were summarized in Excel (Microsoft) as total scores and separately for each document type (initial visit, follow-up, discharge). For completeness, researchers manually reviewed the documents to identify included and omitted items. Perceived accuracy was analyzed by calculating the average scores assigned by clinicians for each document in Excel. We also categorized factors that lowered accuracy based on clinicians' written feedback. To identify these factors, researchers coded the feedback through open coding and conducted a thematic analysis[31].

For qualitative analysis, interviews with both patients and clinicians were transcribed and coded by the researchers (SIK, JYP). Open coding and thematic analysis were conducted to identify themes related to the creation and delivery of the PFCNs, how they facilitated patient participation, and challenges encountered during their use[31]. The coding process was iterative and continued until the researchers reached consensus on the themes. Interviews and surveys were conducted and analyzed in Korean to ensure precise interpretation of participants' original responses.

Results

Participants

A total of 30 participants (10 clinicians and 20 patients) participated in the LLM-generated explanatory document evaluation study [Supplement File 3]. Among the clinician participants, 9 (90%) were junior doctors currently in their residency training. The average age of the clinicians was 29.4 years, with an average work experience of 4.7 years. Five clinicians (50%) were from the Department of Emergency Medicine. The average age of the patient participants was 29.5 years, consisting of 5 men and 15 women. Patients had an average of 4.65 previous ED visits. The primary reasons for their most recent ED visits included trauma (7 cases, e.g., ankle pain from hiking), infections (6 cases, e.g., EBV infection, fever), gastrointestinal problems (5 cases, e.g., abdominal pain, diarrhea), and psychosomatic issues (e.g., dyspnea in stressful situations).

Quality evaluation of PFCNs

General information

In this study, a total of 120 PFCNs (60 from clinicians and 60 from patients) were generated. The examples of the clinical notes written by clinicians and the PFCNs generated by LLM were included in Supplement File 4. The average word count for the 120 clinical notes written by clinicians was 40.52 words, while the average word count for the 120 PFCNs was 231.31 words (Table 2).

Understandability

The PEMAT understandability score for LLM-generated PFCNs was 87.2% for both clinicians and patients, exceeding the 70% threshold and indicating high comprehensibility (Table 2). Among the document types, discharge PFCNs received the highest ratings, while follow-up PFCNs were rated lowest. Notably, patients gave the lowest rating to the item about defining medical terms (Item 4), while clinicians rated a different item regarding the inclusion of information beyond the document's purpose (Item 2) lowest, highlighting differing expectations in content clarity and purpose.

A small-to-moderate positive correlation (Pearson's $r = 0.34$) was observed between patients' post-intervention PPQ scores and understandability, suggesting that even with a high PEMAT understandability score of 87.2%—well above the 70% threshold—document comprehension may only modestly influence patient participation in ED consultations.

Completeness

Among the 60 PFCNs evaluated by clinicians, 55 included all the required elements (Table 2). However, five discharge documents were missing specific content: three instances of missing "History and physical examination details" and two instances of missing "Expected diagnoses."

Perceived accuracy

The average perceived accuracy of the PFCNs was 5.45 out of 6 (6 points: 44 documents, 5 points: 20 documents, 4 points: 5 documents, 3 points: 1 document) (Table 2). The accuracy scores for the three types of documents were: initial consultation document 5.35, follow-up explanation document 5.55, and discharge document 5.45. Since the perceived accuracy scores for the three types of documents did not meet normality assumptions, a Friedman non-parametric repeated measures ANOVA was conducted, which showed no statistically significant difference between the scores ($p = 0.066, > 0.05$).

To address the factors contributing to reduced accuracies of LLM-generated PFCNs, we identified two primary factors: (1) Hallucinations and (2) Misalignments and categorized these two factors into detailed subtypes (Table 3). Hallucinations refer to systematic errors in the content generation process, such as missing, inaccurate, or fabricated information not present in the clinician's original clinical notes[13]. In contrast, misalignments occur when the generated content does not match the clinician's expectations or intentions. This reflects a failure to accurately capture the clinician's reasoning or nuanced communication goals. The key distinction lies in their origins: hallucinations stem from the LLMs' content generation process, while misalignments result from the LLMs' inability to accurately interpret and convey the clinician's intent.

	Phase 1 : clinician's evaluation				Phase 2 : patient's evaluation				Total
Category	**Initial visit ($n = 20$)**	**Follow-up ($n = 20$)**	**Discharge ($n = 20$)**	**Total ($n = 60$)**	**Initial visit ($n = 20$)**	**Follow-up ($n = 20$)**	**Discharge ($n = 20$)**	**Total ($n = 60$)**	**Total ($n = 120$)**
Word count									
Clinical note [Avg. (SD)]	43.8 (13.98)	24.8 (10.77)	29.6 (7.62)	32.73 (13.72)	73.85 (17.02)	31.65 (17.02)	39.4 (7.89)	48.3 (19.83)	40.52 (18.69)
PFCN [Avg. (SD)]	261.5 (10.87)	225 (25.08)	210.5 (22.61)	232.33 (29.66)	264.75 (30.67)	217.6 (24.31)	208.5 (16,91)	230.28 (34.84)	231.31 (32.22)
Understandability									
PEMAT Understandability score [%, Avg. (SD)]	86.7 (8.21)	86.3 (5.87)	88.8 (5.87)	87.2 (6.59)	86.7 (9.07)	86.3 (8.76)	88.8 (9.92)	87.2 (9.13)	87.2 (7.93)
Completeness									
Number of omitted information	0	0	5	5	N/A	N/A	N/A	N/A	N/A
Perceived accuracy									
Total [avg.(SD)]	5.35 (0.75)	5.55 (0.83)	5.45 (0.60)	5.45 (0.72)	N/A	N/A	N/A	N/A	N/A
Usefulness									
USE- Cognition [Avg.(SD)]	N/A	N/A	N/A	N/A	25.2 (3.53)	25.6 (7.62)	24.2 (7.13)	25.0 (5.14)	N/A
USE - Emotion [Avg.(SD)]	N/A	N/A	N/A	N/A	20.8 (7.81)	20.8 (7.63)	21.1 (7.35)	20.9 (7.47)	N/A
USE - Behavior [Avg.(SD)]	N/A	N/A	N/A	N/A	24.1 (5.20)	24.1 (6.79)	24. (7.42)1	24.1 (6.42)	N/A
USE - Total [Avg.(SD)]	N/A	N/A	N/A	N/A	70.0 (14.6)	70.4 (16.9)	69.3 (20.6)	69.9 (17.3)	N/A

Table 2. Quality assessment results of the PFCNs.

Factors	Category	Examples
Hallucination		
	Omission of information	- Outpatient follow-up instruction recorded in the medical note was not reflected in the explanatory document - Expected diagnosis missing from the discharge explanatory document.
	Generation of inaccurate information	- The location of the appendix was incorrectly recorded as being on the left side. - The explanatory document informed the patient that the medication for controlling abdominal pain intended for discharge was already administered in the ED.
	Addition of the unmentioned information	- An abdominal ultrasound test was generated despite not being recorded by the clinician. - The follow-up explanatory note includes information indicating that the patient is eligible for discharge.
Misalignment		
	Lack of reflection of the clinician's clinical reasoning process and routine practices	- Although the reason for empirical antibiotic prescription was 'infection source identification' in the document, the clinician's intent was symptom relief and bacterial eradication - The explanatory document includes factors that may contribute to herniated intervertebral disc (HIVD) (e.g., severe impact from an accident); however, these seem unrelated to the information confirmed by the clinician during the patient's consultation (e.g., a previous history of HIVD, desk job).
	Unwanted disclosure of information	- The explanatory document disclosed the reason why the patient refused an MRI, which was recorded by the clinician for legal protection.
	Awkward phrasing in the explanation	- Inconsistent or unusual patient address forms (e.g., 'Mr. Kim Young-ho,' 'Patient Kim Young-ho')
	Anticipation to the supplement information on clinician's explanation	- Waiting times for blood tests and imaging (CT, MRI) examinations. - Preparation requirements for imaging tests (e.g., removing metallic objects from clothing, remaining still during the procedure).

Table 3. Factors contributing to lower perceived accuracy of the LLM-generated PFCNs.

Category	Before Role-play [Avg. (SD)] ($n=20$)	After Role-Play [Avg. (SD)] ($n=20$)	p-value
PPQ score			
Involvement	9.3 (2.18)	13.35 (2.64)	$p<0.001$
Information	10.3 (1.89)	14.1(1.86)	$p<0.001$
Communication	7.35 (1.31)	10.3 (1.78)	$p<0.001$
Relationship to staff	11.9 (1.94)	16.8 (1.90)	$p<0.001$
Overall assessment of involvement	1.15 (0.37)	1.9 (0.31)	$p<0.001$
Total	40 (6.13)	56.6 (8.56)	$p<0.001$

Table 4. PPQ scores of the patients before and after simulated role-play in ED consultations.

Usefulness

The results of the patients' usefulness evaluation showed USE scores of 70.0 for the initial visit document, 70.4 for the follow-up document, and 69.3 for the discharge document (Table 2). Since the usefulness scores for the three document types did not meet normality assumptions, a Friedman non-parametric repeated measures ANOVA was conducted, which confirmed that the score differences among the initial, progress, and discharge documents were not statistically significant ($p>0.05$).

The usefulness evaluation revealed that patients found PFCNs particularly helpful for acquiring information (Cognition: 8.32), supporting health behaviors (Behavior: 8.02), and providing emotional reassurance (Emotion: 6.96). To explore how perceived usefulness relates to patient engagement, we analyzed the correlation between USE scores and post-study PPQ scores, which better capture the influence of PFCNs on participation. A strong positive correlation (Pearson's $r=0.86$) was observed, suggesting that patients who found the PFCNs more useful were more likely to actively participate in ED consultations.

Patient participation with the use of PFCNs in ED consultations

Changes in PPQ score before and after ED consultation Role-Play

The results showed that the patient participation score in the simulated consultation with PFCNs was 56.5, a statistically significant increase compared to the score of 40 from previous ED consultations ($P<0.001$) (Table 4). Since the subcategories Involvement, Information, and Communication did not meet normality assumptions, the Wilcoxon test was used, and all three showed significant score increases after the simulated consultations ($p<0.001$). The subcategory Relationship to Staff met the normality assumption and was analyzed with a Student's paired samples t-test, which also indicated a significant score increase ($p<0.001$). For the Overall Assessment of Involvement, a Wilcoxon test was conducted, and the score increase was again statistically significant ($p<0.001$).

Supporting patient participation in ED settings through PFCNs

Through the patient interview results, we found that patients perceived PFCNs as supportive documents to actively participate in ED consultations. First, PFCNs encouraged patients to participate more actively in their ED consultations by helping them prepare questions in advance based on the information of the documents (e.g. "Could you let me know if hospitalization will be necessary?") Most patients noted that the hectic ED environment often made it difficult to think of questions during consultations. However, the PFCNs provided

them with time to reflect and formulate more specific questions, with some even revising their questions based on the information in the documents.

> *"I feel like I can prepare more refined questions… The doctor would likely see that I'm engaged and want to participate in decision-making."* (Patient 4).

Second, the PFCNs supplemented the clinician's verbal explanations, helping patients better understand their condition and treatment progress. Especially in Korea's busy tertiary hospitals[32], these documents could be useful for acquiring information about expected diagnosis, test types, test results and discharge education. Patients also reported that during waiting times, they could read the document and look up unfamiliar terms (e.g. appendicitis, migraine) using online search engines to better understand their situation.

> *"With everything organized in the document, I feel like I could've said, 'Hey, I haven't done this yet,' more easily."* (Patient 3).

Third, patients described the PFCNs as "made just for me" and "carefully tailored" to their own situation, which enhanced their trust in the personalized and contextualized information provided. Seeing their own words and the clinician's explanations reflected in the document reassured them that their concerns had been heard and addressed by the clinicians.

> *"I sometimes worry whether the doctor truly understood my situation and symptoms, or if there might have been some mix-up with another patient. But when I saw this document, I realized that what the doctor had said to me was reflected exactly as it was, and that my situation had been accurately assessed. Being able to see that in writing made me trust it more."* (Patient 5).

Fourth, patients reported that reviewing the PFCNs provided emotional reassurance, particularly in the busy and often uncomfortable ED environment where essential information (e.g., details about the ED care process, upcoming treatment plans) was frequently lacking. Revisiting the clinician's explanations helped alleviate the anxiety they felt during the consultation.

> *"Unless it's an extremely critical situation, most ED patients are conscious and aware. Seeing the explanations laid out in the document helps me feel reassured. I can clearly see my condition and the steps ahead, which makes me feel more at ease."* (Patient 9).

Finally, patients reported that the PFCNs could improve their relationships with clinicians. By reducing unnecessary questions during consultations, patients perceived that the clinician could focus more on their specific needs and concerns. Clinicians also reported that receiving pre-submitted patient questions allowed them to better understand the patient's specific concerns, level of comprehension, and health literacy, enabling them to adjust their communication style and explanations to better meet the patient's needs. As a result, clinicians anticipated that the PFCNs could ease the burden of challenging conversations and enhance the overall quality of patient interactions.

> *"I feel like this would make the doctor-patient relationship more positive because I wouldn't have that unnecessary guilt for asking too many questions. From the doctor's perspective, having pre-prepared questions would probably be much easier than answering impromptu ones. It could even improve workflow efficiency."* (Patient 2).

Challenges in the applying PFCNs to ED consultations

Despite the benefits, several challenges emerged, particularly from the clinicians' perspective. One major concern was the risk of misinformation from the hallucination and misalignments of the PFCNs. Clinicians highlighted cases where the PFCNs included inaccurate or extraneous details not found in the original clinical notes—such as test results that were not conducted—which led to patient confusion and complaints. This discrepancy not only increased the clinicians' communication burden but also raised concerns about legal liability.

> *"The ultrasound inclusion seemed like a system error. If there's a mismatch between tests conducted and those mentioned, patients often file complaints, and it can become a big issue. "* (Clinician 9).

Another concern was the mismatch between patients' and clinicians' expectations of the document's content. While patients expressed a desire for more comprehensive explanations (e.g., test preparation, discharge education), clinicians preferred to keep the content limited to what was explicitly documented, to avoid legal and ethical risks. This divergence created tension between the perceived utility and safety of the PFCNs.

> *" I think more detailed information could have been included. For example, during the long waiting times, I could've read about what a CT scan involves and how to prepare for it. Or there could've been more detailed post-discharge instructions, like dietary recommendations."* (Patient 7).

We also identified that patients tended to accept the documents uncritically if they aligned with the verbal explanation, even when errors were present. This reliance on perceived consistency rather than factual accuracy could lead to false reassurance or misunderstanding, especially in cases where nuanced clinical intent was not properly captured in the automatically generated content.

> *"Since the content matched exactly what the doctor told me, I just accepted it as is. I didn't feel there was anything untrustworthy about it."* (Patient 1).

Discussion

This study identified how LLM-generated PFCNs could support patient participation in ED consultations. Our results showed that PFCNs improved not only patients' perceived involvement in their care but also addressed key challenges in the ED environment. In line with the PEMAT understandability score of 87.2% (> 70%), patients reported that the documents helped them understand their condition, prepare more targeted questions, and feel emotionally reassured—outcomes that are especially meaningful in the fast-paced, high-pressure ED setting. These improvements were reflected in significantly higher PPQ scores ($P < 0.05$) and qualitative feedback describing more active communication and stronger relationships with clinicians. Although this study did not directly compare clinician-written and LLM-generated documents, the observed understandability and accuracy scores were similar to those reported in previous studies[13,33], reinforcing the feasibility of using PFCNs as a scalable and patient-centered solution in ED settings.

Since providing PFCNs during ED consultations is a new concept, unlike previous studies[13,15,33], we did not conduct direct comparisons between clinician-written and LLM-generated documents. Nevertheless, the quality of the LLM-generated PFCNs in our study was comparable to prior work. For example, Zaretsky et al. (2024) reported a PEMAT understandability score of 81% for GPT-4–generated inpatient discharge summaries, whereas our PFCNs achieved 87.2%. In terms of accuracy, their study found that 54% of documents received the top rating, compared with 73.3% in ours. These findings support the potential of LLMs to produce clinical documents at a quality level similar to or exceeding that of clinician-written materials. Importantly, our work moves beyond quality assessment alone by examining the real-world experiences and perceptions of patients in the high-pressure, time-constrained ED environment. By examining patients' perceptions and experiences at moments when they encountered confusion or sought additional information while reviewing LLM-generated documents, our study offers ED-specific, patient-centered insights not addressed in prior research. These insights underscore not only the technical feasibility of producing high-quality PFCNs, but also their practical value in enhancing comprehension, encouraging active participation, and guiding the design of more personalized, context-sensitive patient materials for future ED care.

The study results identify that PFCNs can support both patients and clinicians by addressing key information gaps that hinder patient-centered care in the ED. Prior research has highlighted several limitations of paper-based documents such as printed discharge instructions or education materials—most notably, their static nature, the frequent use of complex medical terminology, and the lack of timely personalization during the care process[34–36]. Addressing these challenges, LLM-generated documents can be rapidly tailored to the patient's current condition, integrate plain language explanations, and be produced on demand during ED consultations—advantages that are particularly valuable in a time-constrained, high-pressure environment. In this study, PFCNs reflected the content of clinician–patient consultations, enabling the delivery of personalized, contextually relevant information. This not only reinforced patients' trust that clinicians had accurately understood their symptoms and circumstances but also provided key information at each stage of the consultation (e.g., treatment plans, probable diagnoses), offering emotional reassurance. These findings highlight the potential of LLMs to meet immediate informational needs and adapt to frequent changes in the care process in the ED, where each patient's symptoms and required information vary substantially.

However, unlike printed patient education reports, LLM-generated documents also present new challenges, including the need for robust quality control to prevent hallucinated or misleading content, as well as considerations for medico-legal liability. These challenges underscore the importance of implementing robust safeguards within clinician–patient communication to ensure that LLM-generated documents truly enhance patient understanding and engagement. To ensure that PFCNs can be safely and efficiently integrated into ED workflows, further research is needed to develop strategies for quality assurance and risk mitigation.

Based on the study results, we suggest a set of practical implement strategies and future research directions to guide the safe and effective adoption of PFCNs in real-world ED settings:

(1) Clinician-in-the-Loop Validation after Document Generation : Given the risks of hallucinations and misalignments identified in our study, we recommend adopting a Clinician-in-the-Loop (CITL) framework[37], where clinicians can efficiently oversee and approve PFCNs with minimal additional burden. Future research should explore how LLM-based system design can better support clinicians in identifying and correcting errors in these documents. For instance, the system could automatically flag high-risk content—such as diagnoses and medications—for mandatory review, while allowing lower-risk educational content to be optionally reviewed. Considering prior research proposed a streamlined EHR interface, MedKnowts[38] by automatically displaying relevant text snippets for labs, tests, and other items, the systems might integrate inline editing tools—such as one-click deletion or summarization toggles—to enable efficient review and refinement without requiring clinicians to rewrite the entire document.

(2) Transparent Information Attribution at Document Delivery: To address the tension between patients' desire for detailed explanations and clinicians' concerns about miscommunication and liability, a transparent information attribution can be embedded in the PFCN delivery stage. For example, based on our study results, clinically crucial informations (e.g. diagnoses, test plans) can be labeled as "clinician-authored," while educational content (e.g., "What is a CT scan?") can be marked as "AI-generated." Additionally, a system-generated disclaimer can be displayed at the top or bottom of each document, stating that portions of the content were generated by AI and may require confirmation by clinicians. Future research is needed to explore how visual cues can be designed and applied to support patients intuitively navigate the source and reliability of information.

(3) Post-Consultation Review via PHR Systems : A common feature of ED workflows is prolonged waiting periods between triage, diagnostic testing, and follow-up consultations[39,40]. Given that many hospitals

increasingly adopt personal health record (PHR) systems to deliver patient clinical information, these intervals offer an ideal opportunity to deliver LLM-generated PFCNs to patients via mobile health platforms (e.g., MyChart[41], HealtheLife[42]. Our results showed that the question feature could promote patients to understand the content and communicate with clinicians. Therefore, the future systems could incorporate features such as interactive Q&A tools, annotation or highlighting capabilities, and links to institutionally approved educational content.

(4) Legal and Ethical Considerations in AI-Assisted Patient–Clinician Communication: Our findings suggest that deploying LLM-generated documents in ED settings may present potential legal and ethical challenges for clinician-patient communication. Legally, inaccuracies—such as listing tests that were not performed—can induce patient confusion, complaints, and even medico-legal claims, underscoring the need for clearly defined legal responsibility in AI-generated documentation. Additionally, patient questions or comments submitted via the system may be interpreted as formal medical intent; failure to review or address them may carry liability implications. Therefore, adherence to regulatory frameworks such as the Health Insurance Portability and Accountability Act (HIPAA) and the General Data Protection Regulation (GDPR) is essential. Explicit guidelines should be established regarding informed consent mechanisms, authorship attribution, audit trails, and the secure long-term storage of AI-generated communications.

Ethically, the results showed the potential risk of overreliance, as patients may accept AI-generated content uncritically when it aligns with clinicians' verbal explanations—even if errors remain after clinician validation. These tendencies may arise from differing expectations of AI, trust levels, and access to information of patients[43,44]. To reduce this risk, recent studies have proposed strategies such as providing partial explanations to limit unnecessary trust[45] or incorporating cognitive forcing functions[46] that encourage users to think critically—for example, prompting patients to enter their own questions or articulate the reasoning behind their decisions. Therefore, future research should explore how these strategies can be integrated into AI systems and assess, in real patient–clinician communication settings, how they influence patients' reliance on the system and their participation in care.

Despite the promising results, the study has several limitations. First, considering the potential risks of LLM-generated content, this study was conducted as an in-lab simulation using limited scenarios, which did not fully reflect the diverse cases, situations, and patient populations encountered in real-world ED. Therefore, the high urgency, noise, and extreme anxiety and pain experienced by patients in a real ED differ from the cognitive capacity assumed in this simulation, which could negatively impact the acceptance and utility of PFCNs.

Second, the generalizability of findings is constrained by the small sample size and the demographic homogeneity of participants, who were predominantly young, female, and digitally literate individuals. Although patient participation (PPQ) statistically significantly increased ($P < 0.001$), it must be considered that these positive results may not be directly reproducible in a chaotic real-world ED environment, or with older patients or those with low health literacy.

Third, the study used only one LLM (GPT-4o), without comparing outputs from other widely used models such as Claude, LLaMA, and Gemini. Furthermore, we clarified that no comparison was made between the AI-generated summaries and plain-language summaries written by medical professionals (human-generated), which limits our ability to demonstrate the unique value or efficiency of using an LLM.

Finally, due to the lack of validated tools for assessing Korean-language AI-generated clinical documents, especially in ED settings, future research should develop and validate context-specific instruments to evaluate accuracy, completeness, and patient participation.

Conclusions

We evaluated the potential of LLM-generated PFCNs to enhance patient participation in emergency care. This study demonstrates the feasibility of using LLMs to transform clinician notes into patient-friendly documents that improve understanding, communication, and relationships during ED consultations. Our findings also highlight key challenges, including hallucinations and alignment issues, underscoring the need for careful design and oversight. These results suggest that with thoughtful integration, PFCNs can serve as a valuable tool to promote patient-centered care in real-world ED settings.

Data availability

The data sets generated during this study are available from the corresponding author upon reasonable request.

Received: 10 April 2025; Accepted: 29 November 2025
Published online: 05 December 2025

References

1. Walsh, A. et al. Patient-centered care in the emergency department: a systematic review and meta-ethnographic synthesis. *Int. J. Emerg. Med.* **15**, 36 (2022).
2. Berg et al. Questionnaire measuring patient participation in health care: scale development and psychometric evaluation. *Eur. J. Cardiovasc. Nurs.* **19**, 600–608 (2020).
3. Wang, L. et al. The emergency patient's participation in medical decision-making. *J. Clin. Nurs.* **25**, 2550–2558 (2016).
4. Frank, C., Fridlund, B., Baigi, A. & Asp, M. Patient participation in the emergency department: an evaluation using a specific instrument to measure patient participation (PPED). *J. Adv. Nurs.* **67**, 728–735 (2011).
5. Aung, E., Donald, M., Williams, G. M., Coll, J. R. & Doi, S. A. R. Joint influence of patient-assessed chronic illness care and patient activation on glycaemic control in type 2 diabetes. *Int. J. Qual. Heal Care.* **27**, 117–124 (2015).

6. Sharma, A. E. et al. Patient engagement in health care safety: an overview of Mixed-Quality evidence. *Heal Aff.* **37**, 1813–1820 (2018).
7. Gillani, S. M. R., Nevill, A. M. & Singh, B. M. A randomised controlled trail in diabetes demonstrating the positive impact of a patient activation strategy on diabetes processes and HbA1c: the WICKED project. *Br. J. Diabetes*. **17**, 58–63 (2017).
8. Sacristán, J. A. Patient-centered medicine and patient-oriented research: improving health outcomes for individual patients. *BMC Méd Inf. Decis. Mak.* **13**, 6 (2013).
9. Lateef, F. Patient expectations and the paradigm shift of care in emergency medicine. *J. Emergencies Trauma. Shock*. **4**, 163–167 (2011).
10. Park, S. Y. & Chen, Y. Design Opportunities for Supporting Patient Information Needs during an Emergency Visit. *Proc. 10th EAI Int. Conf. Pervasive Comput. Technol. Healthc.* https://doi.org/10.4108/eai.16-5-2016.2263339 (2016).
11. Johnson, C., Richwine, C. & Patel, V. Individuals' Access and Use of Patient Portals and Smartphone Health Apps, 2020. ASTP Health IT Data Brief [Internet] (2020).
12. Xiao, Z., Li, T. W., Karahalios, K. & Sundaram, H. Inform the uninformed: improving online informed consent reading with an AI-Powered chatbot. *ArXiv* https://doi.org/10.48550/arxiv.2302.00832 (2023).
13. Zaretsky, J. et al. Generative artificial intelligence to transform inpatient discharge summaries to Patient-Friendly Language and format. *JAMA Netw. Open.* **7**, e240357 (2024).
14. Bala, S., Keniston, A. & Burden, M. Patient perception of Plain-Language medical notes generated using artificial intelligence software: pilot Mixed-Methods study. *JMIR Form. Res.* **4**, e16670 (2020).
15. Decker, H. et al. Large Language Model–Based chatbot vs Surgeon-Generated informed consent Documentation for common procedures. *JAMA Netw. Open.* **6**, e2336997 (2023).
16. Asgari, E. et al. A framework to assess clinical safety and hallucination rates of LLMs for medical text summarisation. *Npj Digit. Med.* **8**, 274 (2025).
17. Shah, S. V., Accuracy & Consistency Hallucination of large Language models when analyzing unstructured clinical notes in electronic medical records. *JAMA Netw. Open.* **7**, e2425953 (2024).
18. Roustan, D. & Bastardot, F. The clinicians' guide to large Language models: A general perspective with a focus on hallucinations. *Interact. J. Méd Res.* **14**, e59823 (2025).
19. Das, A. B., Ahmed, S. & Sakib, S. K. Hallucinations and key information extraction in medical texts: A comprehensive assessment of Open-Source large Language models. *ArXiv* https://doi.org/10.48550/arxiv.2504.19061 (2025).
20. Subramanian, C. R., Yang, D. A. & Khanna, R. Enhancing health care communication with large Language Models—The Role, Challenges, and future directions. *JAMA Netw. Open.* **7**, e240347 (2024).
21. Tjan, T. E., Wong, L. Y. & Rixon, A. Conflict in emergency medicine: A systematic review. *Acad. Emerg. Med.* **31**, 538–546 (2024).
22. Greenwood, M. J. & Heninger, J. R. Structured communication for patient safety in emergency medical services: A legal case report. *Prehospital Emerg. Care*. **14**, 345–348 (2010).
23. Du, Z. et al. LLMs can simulate standardized patients via agent Coevolution. *ArXiv* https://doi.org/10.48550/arxiv.2412.11716 (2024).
24. Louie, R. et al. Roleplay-doh: enabling Domain-Experts to create LLM-simulated patients via eliciting and adhering to principles. *ArXiv* https://doi.org/10.48550/arxiv.2407.00870 (2024).
25. Inc., V. & Evertime https://everytime.kr/ (2024).
26. OpenAI. GPT-4o. https://openai.com/
27. Vercel Next.js. https://nextjs.org/ (2025).
28. Shoemaker, S. J., Wolf, M. S. & Brach, C. Development of the patient education materials assessment tool (PEMAT): A new measure of understandability and actionability for print and audiovisual patient information. *Patient Educ. Couns.* **96**, 395–403 (2014).
29. Hölzel, L. P. et al. Usefulness scale for patient information material (USE) - development and psychometric properties. *BMC Méd Inf. Decis. Mak.* **15**, 34 (2015).
30. Project, T. J. Jamovi. https://www.jamovi.org (2024).
31. Braun, V. & Clarke, V. Using thematic analysis in psychology. *Qualitative Res. Psychol.* **3**, 77–101 (2006).
32. Chang, H. et al. Emergency department crowding: a National data report. *Clin. Exp. Emerg. Med.* **11**, 331–334 (2024).
33. Tung, J. Y. M. et al. Comparison of the quality of discharge letters written by large Language models and junior clinicians: Single-Blinded study. *J. Méd Internet Res.* **26**, e57721 (2024).
34. Bhattad, P. B. & Pacifico, L. Empowering patients: promoting patient education and health literacy. *Cureus* **14**, e27336 (2022).
35. Stossel, L. M., Segar, N., Gliatto, P., Fallar, R. & Karani, R. Readability of patient education materials available at the point of care. *J. Gen. Intern. Med.* **27**, 1165–1170 (2012).
36. Fage-Butler, A. M. & Jensen, M. N. Medical terminology in online patient–patient communication: evidence of high health literacy? Heal. *Expect* **19**, 643–653 (2016).
37. Tang, S., Modi, A., Sjoding, M. W. & Wiens, J. Clinician-in-the-Loop Decision Making: Reinforcement Learning with Near-Optimal Set-Valued Policies. arXiv https://doi.org/10.48550/arxiv.2007.12678 (2020).
38. Murray, L. et al. MedKnowts: unified Documentation and information retrieval for electronic health records. *34th Annu. ACM Symp. User Interface Softw. Technol.* **1169-1183** https://doi.org/10.1145/3472749.3474814 (2021).
39. van der Veen, D., Remeijer, C., Fogteloo, A. J., Heringhaus, C. & de Groot, B. Independent determinants of prolonged emergency department length of stay in a tertiary care centre: a prospective cohort study. *Scand. J. Trauma. Resusc. Emerg. Med.* **26**, 81 (2018).
40. Payne, K., Risi, D., O'Hare, A., Binks, S. & Curtis, K. Factors that contribute to patient length of stay in the emergency department: A time in motion observational study. *Australas Emerg. Care*. **26**, 321–325 (2023).
41. Redelmeier, D. A. & Kraus, N. C. Patterns in patient access and utilization of online medical records: analysis of mychart. *J. Méd Internet Res.* **20**, e43 (2018).
42. Porter, A. S., O'Callaghan, J., Englund, K. A., Lorenz, R. R. & Kodish, E. Problems with the problem list: challenges of transparency in an era of patient curation. *J. Am. Méd Inf. Assoc.* **27**, 981–984 (2020).
43. Shekar, S., Pataranutaporn, P., Sarabu, C., Cecchi, G. A. & Maes, P. People over trust AI-generated medical responses and view them to be as valid as doctors, despite low accuracy. *ArXiv* https://doi.org/10.48550/arxiv.2408.15266 (2024).
44. Mesinovic, M., Watkinson, P. & Zhu, T. Explainability in the age of large Language models for healthcare. *Commun. Eng.* **4**, 128 (2025).
45. Vasconcelos, H. et al. Explanations Can Reduce Overreliance on AI Systems During Decision-Making. *Proc. ACM Hum.-Comput. Interact.* **7**, 1–38 (2023).
46. Buçinca, Z., Malaya, M. B. & Gajos, K. Z. To trust or to think. *Proc. Acm Hum Comput Interact.* **5**, 1–21 (2021).

Author contributions

Sung-In Kim and Joonyoung Park conducted the studies, analyzed the data, and contributed to writing the manuscript. Taewan Kim and Wooseok Seo designed the interviews and user studies, as well as provided feedback on the manuscript. Tae Rim Kim and Won Chul Cha reviewed the study goal, advised how to properly implement LLMs to generate patient friendly clinical notes in ED settings and supported participant recruitment at Sam-

sung Medical Center. Hwajung Hong, as the corresponding author, supervised the overall study implementation and analysis processes. All authors reviewed and approved the final manuscript.

Funding
This research was supported by National Research Foundation of Korea (NRF) funded by the Korea government (MSIT) (No. RS-2023-00262527).

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-31113-y.

Correspondence and requests for materials should be addressed to H.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.