scientific reports

OPEN

Shifts in emergency physicians' attitudes toward large language model-based documentation: a pre- and post-implementation study

Seongwon Lee[1,5], Ji Woo Song[2,5], Seng Chan You[1,3]✉ & Ji Hoon Kim[1,4]✉

Large language models (LLMs) can assist physicians in writing medical notes more efficiently. This study evaluates whether using an LLM assistant for writing emergency department discharge notes can reduce doctors' workload and addresses concerns regarding the incorporation of AI in medical practice. Eight emergency doctors with an average experience of 12 years participated in our study. We surveyed them prior to, post 3 days, and post 5 weeks of their LLM usage. The results showed that doctors' concerns about using LLMs decreased significantly and remained low throughout the study period. Moreover, the LLM usage considerably reduced the perceived workload, with the time required to write each discharge note reduced by one-third of the original time. These findings demonstrate that doctors readily accept and benefit from LLM assistants in their daily practice. Our study provides the first real-world evidence of how doctors' attitudes toward AI assistants change over time in clinical settings, offering valuable insights into the future implementation of LLM-based documentation tools in healthcare.

Keywords Surveys and questionnaires, Natural language processing, Workload, Emergency service, hospital, Longitudinal studies

Large language models (LLMs) have attracted considerable attention as a promising technology for clinical documentation owing to their ability to rapidly assimilate, summarize, and rephrase information[1–5]. An LLM-based clinical documentation system is expected to enhance the efficiency of medical note writing and reduce healthcare professionals' administrative burden[6]. Numerous studies have elaborated on the development and evaluation of LLM-based clinical documentation technologies[6–9], but limited studies have empirically investigated the effects of the LLM system on healthcare professionals' practice and their attitudes toward the system in real-world clinical practice.

According to a survey by the American Medical Association in 2024, 40% of 1,183 physicians expressed equal excitement and concerns about the increased use of AI in healthcare[10]. This ambivalence highlights the critical role of physicians' acceptance and attitudes in the successful integration of AI-based systems. Prior studies have emphasized that trust and acceptance among healthcare professionals are indispensable prerequisites for the widespread implementation of AI in clinical practice[11].

We conducted a longitudinal survey study on the "Your-Knowledgeable Navigator of Treatment-Emergency department Discharge Note assistant" (Y-KNOT-EDN)—an on-premise LLM-based clinical documentation system at Severance Hospital (Seoul, South Korea) implemented in November 2024; this system generates discharge note drafts from electronic health record (EHR) data for physician review and finalization[12]. The Y-KNOT-EDN is a fully EHR-integrated, on-premise AI assistant developed based on the emergency department (ED) discharge note interface at desktop clinical workstations. Physicians trigger drafting from within the EHR by selecting the "AI Note" function, which then generates a draft within seconds from structured and unstructured EHR data. The draft is inserted into the discharge note field for physician review, modification, and sign-off to ensure full clinician oversight at every documentation stage (see Supplementary Figure S1).

[1]Yonsei Institute for Digital Health, Yonsei University, Seoul, Korea. [2]Yonsei University College of Medicine, Seoul, Korea. [3]Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, South Korea. [4]Department of Emergency Medicine, Yonsei University College of Medicine, Seoul, Korea. [5]These authors contributed equally: Seongwon Lee and Ji Woo Song. ✉email: chandryou@yuhs.ac; JICHOON81@yuhs.ac

We aimed to investigate physicians' perceptions and acceptance of LLM-based clinical documentation systems in clinical settings, focusing on ED discharge notes. Specifically, we examined changes in perceived workload and concerns related to LLM use, as well as key technology acceptance factors, including perceived usefulness, attitude, intention to use, and intention to delegate drafting tasks to the LLM assistant system.

Three surveys were administered to emergency physicians responsible for documenting discharge notes at three time points: prior to Y-KNOT-DEN implementation (T1), 3 d after implementation (T2), and 5 weeks after implementation (T3) (see Fig. 1).

Results

Eight emergency medicine attending physicians (four men and four women; mean age 39.9 years) with an average of 12.1 years of ED experience participated in the surveys. On average, the participants handled 10.9 discharge notes daily.

Figure 2 shows the responses to eight major concerns regarding the use of LLMs in discharge note writing over time. Overall concern, defined as the mean across dimensions, reduced from 3.4 at T1 to 2.7 at T2 and further to 2.5 at T3 ($p = 0.008$), showing a 26% reduction from T1 to T3. Pairwise comparisons indicated reductions from T1 to T2 ($p = 0.023$) and from T1 to T3 ($p = 0.008$), whereas T2 and T3 were comparable ($p = 0.202$).

Four dimensions of concern regarding LLMs—worsening patient care, loss of control, generating impersonal drafts, and legal and ethical issues—decreased significantly over time ($p = 0.004$, 0.002, 0.010, and 0.028, respectively). The other four dimensions of concern—false information, data bias, privacy, and worsening physician reasoning—declined without statistical significance ($p = 0.089$, 0.268, 0.143, and 0.156, respectively). These findings support our first hypothesis that physicians' concerns regarding the use of LLMs in clinical documentation would be alleviated after implementation, as overall concern scores significantly decreased from T1 to T2 and T3. Detailed statistics are provided in Supplementary Tables S1 and S2.

Figure 3 shows the perceived workload scores across dimensions. The overall workload dropped from 11.0 at T1 to 8.0 at T2 and further to 6.9 at T3 ($p = 0.040$), showing a 37% reduction from T1 to T3. Paired comparisons indicated comparable reductions between T1 and T2 ($p = 0.023$), T1 and T3 ($p = 0.035$); and T2 and T3 ($p = 0.402$).

The two dimensions of perceived workload—temporal demand and effort required for accomplishment—significantly decreased over time ($p = 0.002$ and 0.021, respectively). The three dimensions of perceived workload—mental demand, physical demand, and frustration—consistently declined without statistical significance ($p = 0.381$, 0.409, and 0.122, respectively), whereas dissatisfaction with performance remained consistently low. Consistent with our second hypothesis, the perceived workload in drafting discharge notes markedly decreased over time, particularly in temporal demand and effort, demonstrating the system's capacity to reduce administrative burden. The detailed statistics are provided in Supplementary Tables S3 and S4.

Participants reported that completing a single discharge note manually required 127.5 s at T1, whereas it required 42.8 s using the LLM assistant at T3, indicating a time reduction of approximately two-thirds ($p = 0.002$).

The participants' perceptions of the LLM assistant at T2 and T3 are shown in Supplementary Figure S1. The mean perceived usefulness of the assistant was 3.8 at T2, which increased to 4.2 at T3 ($p = 0.038$). Attitudes

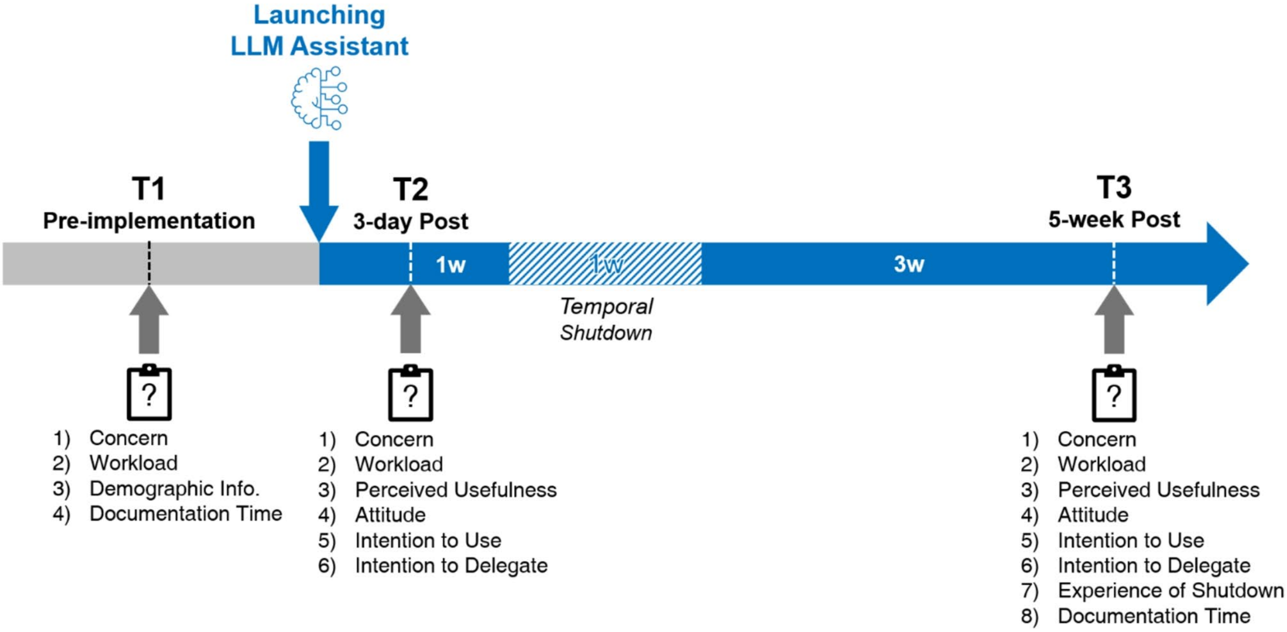

Fig. 1. Timeline of the study and items asked at each phase. The study timeline illustrates three survey phases: T1 (before the Y-KNOT-EDN implementation), T2 (3 days after implementation), and T3 (5 weeks after implementation). Between T2 and T3, the Y-KNOT-EDN was shut down unexpectedly and restored within a week. The survey measures conducted in each phase are indicated below the timeline.

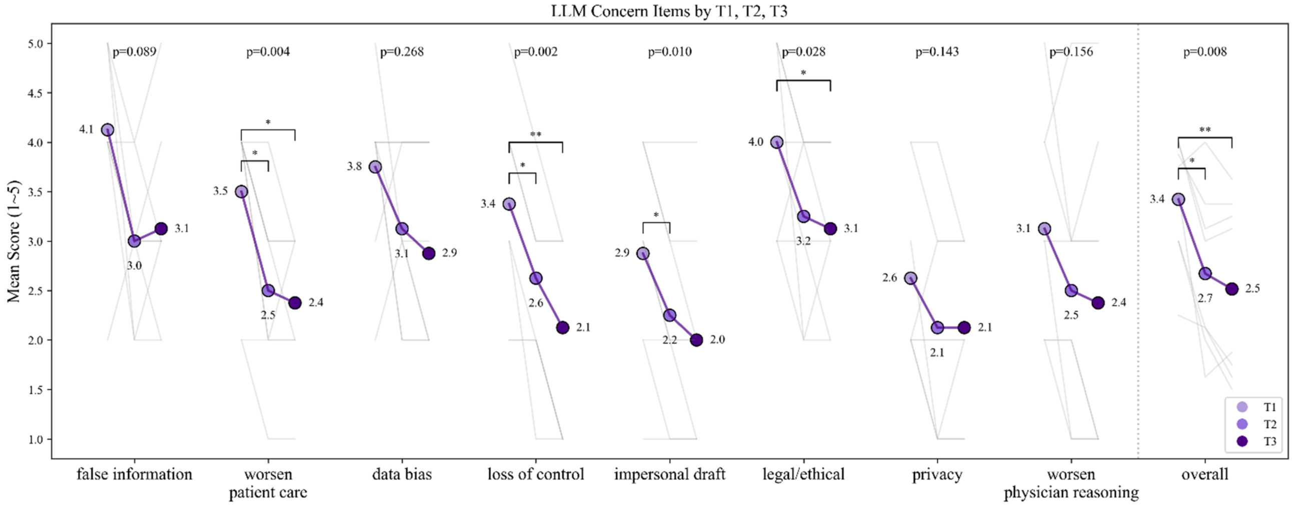

Fig. 2. Concerns on using LLM in clinical documentation over time (T1, T2, and T3). The mean scores for the eight major concerns regarding the use of the LLM-based documentation system, as well as overall concerns, are shown across T1, T2, and T3. Gray lines connect each physician's individual trajectory; large purple circles and solid lines denote group means at T1, T2, and T3. Significant reductions were observed in overall concerns ($p = 0.008$), particularly worsening patient care ($p = 0.004$), loss of control ($p = 0.002$), impersonal drafts ($p = 0.010$), and legal/ethical issues ($p = 0.028$), reflecting increased trust and acceptance of the LLM-based documentation system over time. Note. ***: $p < 0.001$; **: $p < 0.01$; *: $p < 0.05$.

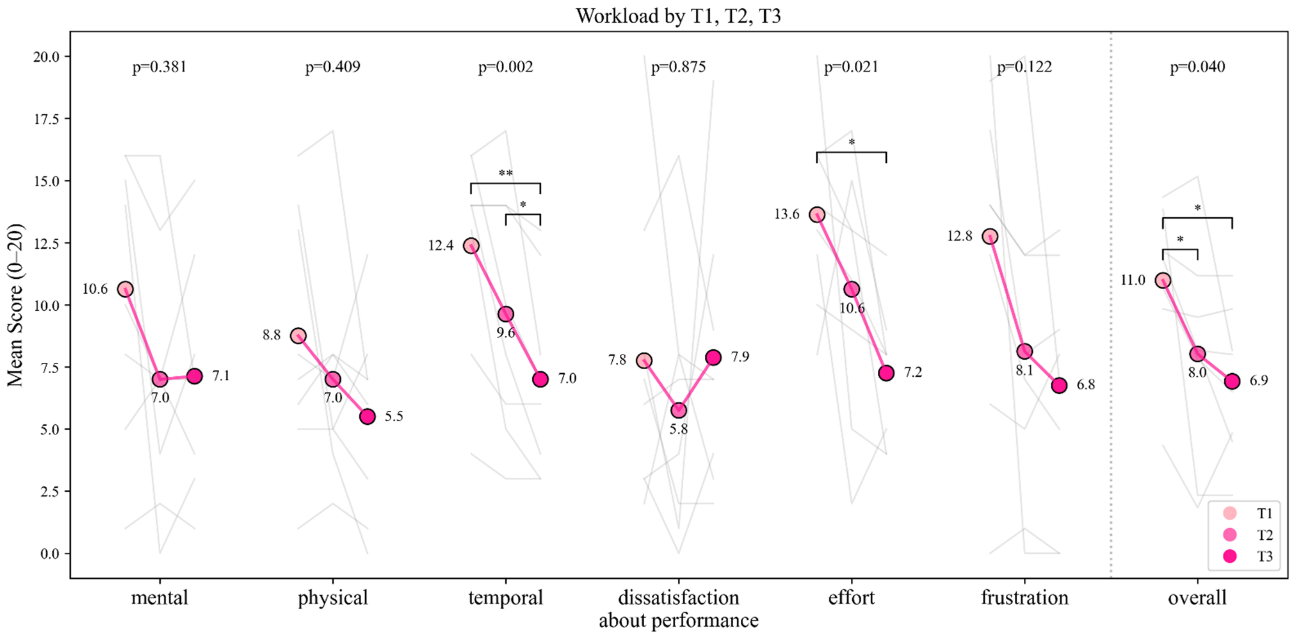

Fig. 3. Perceived workload over time (T1, T2, and T3). The mean NASA-TLX workload scores (0–20 scale) across six sub-dimensions—mental demand, physical demand, temporal demand, dissatisfaction with performance, effort required for accomplishment, and frustration—and the overall workload are shown across T1, T2, and T3. The gray lines show individual physicians' workload trajectories; the large pink circles and connecting lines indicate the group means at each time point. Notable decreases in temporal demand ($p = 0.002$), effort ($p = 0.021$), and overall workload ($p = 0.040$) suggest a reduced perceived burden over the three survey phases. *Note.* ***: $p < 0.001$; **: $p < 0.01$; *: $p < 0.05$.

toward the assistant (T2 mean = 3.7; T3 mean = 3.9), intention to use the assistant (T2 mean = 3.9; T3 mean = 4.0), and intention to delegate discharge drafting to the assistant (T2 mean = 4.0; T3 mean = 4.0) remained consistently high. Regarding the third hypothesis, although acceptance measures remained consistently high from T2 to T3, not all mean scores exceeded 4.0, suggesting partial rather than full support for this expectation.

It is noteworthy that the LLM assistant unexpectedly shut down between T2 and T3, and this was quickly restored within a week. Among the six participants who were affected, three reported that writing discharge notes manually without the aid of the LLM assistant was "Difficult" (score = 4), two found it "Not difficult" (score = 2), and one responded "Not at all difficult" (score = 1), indicating varying levels of perceived disruption.

In response to open-ended questions at T3, physicians most frequently cited reduced documentation time as the primary benefit of the Y-KNOT-EDN, often linking it to decreased workload and improved convenience. When evaluating the generated drafts, participants generally acknowledged high accuracy with fewer hallucinations than expected, but noted instances of irrelevant or redundant content and a lack of patient-specific details, particularly for complex cases. Most respondents agreed that the system had a positive impact on efficiency, particularly by reducing the time spent on nonclinical administrative tasks. For future improvements, physicians requested expanded input data sources (e.g., laboratory results, imaging interpretations, and consultation replies), better handling of complex and multidisciplinary cases to avoid duplication, inclusion of explanatory notes for abnormal results, functionality enhancements such as case-type-specific templates, and more flexible initiation of draft generation. The open-ended questions and physicians' responses are listed in Supplementary Table S5.

Discussion

The key findings of this longitudinal field study are summarized as follows. First, concerns about LLMs were significantly alleviated over time (26% reduction from T1 to T3), particularly those concerning worsening patient care, loss of control, impersonal drafts, and legal and ethical issues. Second, perceived workload declined significantly over time (a 37% reduction from T1 to T3), particularly in terms of temporal demand and effort. The recall-based documentation time was reduced by approximately two-thirds. Moreover, the participants reported high levels of perceived usefulness, attitude, intention to use, and intention to delegate discharge note drafting throughout the study period, indicating that the positive results were not merely driven by a novelty effect but reflected genuine acceptance of the system. Despite these positive outcomes, the open-ended feedback indicates the need for further system refinement to improve accuracy and provide more physician-tailored content. Taken together, these findings supported all three hypotheses proposed in this study, demonstrating that physicians' concerns and perceived workload decreased after implementation while their acceptance of the LLM assistant remained consistently high.

The success of AI implementation in healthcare depends not only on technical performance but also on how healthcare professionals respond to and interact with these systems[11]. Lambert et al. identified the key barriers to healthcare AI adoption, including concerns about the loss of professional autonomy, difficulties with clinical workflow integration, and alert fatigue from oversensitive systems[13]. This study is among the first to examine physicians' perceptions as potential enablers and disablers to the adoption of an LLM-based clinical documentation system, which leverages EHR data extraction and synthesis to automatically generate structured medical notes. By capturing how these perceptions evolved from preimplementation to 5 weeks postimplementation, we provide valuable insights for future AI integration efforts in healthcare. Our findings underscore the necessity of conducting similar pre-post implementation studies when deploying AI systems to understand user experiences, address potential barriers to adoption, and optimize implementation strategies.

The primary goal of implementing an LLM assistant is to reduce physicians' workload[14,15]. However, a previous study reported that AI-powered clinical documentation does not enhance clinician efficiency in primary care settings[16]. Although this study did not measure time using EHRs, our findings demonstrated significant reductions in perceived temporal demand and required effort from preimplementation to 5 weeks postimplementation. Furthermore, aspects less directly related to discharge note writing, such as physical demands and dissatisfaction with performance, showed minimal changes similar to mental demands, suggesting that the cognitive aspects of documentation remained relatively consistent. These results are particularly meaningful for emergency physicians who must manage multiple patients under time constraints because the LLM assistant provided substantial benefits in terms of perceived burden, while preserving the physicians' role in the process. Understanding these nuanced workload changes is critical to guiding the development and acceptance of future AI documentation technologies in healthcare settings.

In addition to workload and acceptance, note quality is a prerequisite for safe adoption. In a preimplementation evaluation of the Y-KNOT-EDN[12], blinded physician ratings of LLM-drafted emergency discharge notes demonstrated high performance across quality domains; for consistency, coherence, fluency, relevance, and safety, it scored 4.78, 4.60, 4.55, 4.72, and 4.73 on a scale of 1–5, respectively, with a subjective satisfaction score of 3.95/5 and usability of 3.32/4, showing that efficiency did not trade off fidelity. Coupled with our longitudinal finding that LLM-related concerns dropped within 2–3 days and remained low at five weeks, this suggests that brief, real-world exposure to a high-quality, EHR-embedded assistant can rapidly convert ambivalence into sustained acceptance. Workload relief, mainly in terms of temporal demand and effort, likely reflects seamless integration into existing workflows rather than added features, underscoring the value of minimizing context switching. Finally, because some determinants of acceptance change after real-world exposure and others remain stable, we emphasize the importance of structured pre- and post-implementation surveys. As demonstrated in our study, this framework offers a systematic way to capture evolving perceptions of AI systems and can be readily adapted for future AI integration research.

Although quantitative survey results suggested strong physician acceptance, the open-ended feedback indicated the need for further improvement in generating more patient-tailored and contextually rich drafts. Several improvement requests, such as integrating broader clinical data (laboratory, imaging, and consultation records) and refining the output for complex or atypical cases, were directly actionable in the next iteration of the Y-KNOT-EDN. Because the current model (Llama-3-8B) has a limited context length and a higher risk of hallucinations, planned upgrades to more advanced models (e.g., MedGemma-27B from Google[17] and gpt-oss-20B from OpenAI[18] combined with expanded and structured reference inputs are expected to address these concerns. This iterative refinement, guided by structured feedback, will enable future evaluations to measure whether satisfaction with and trust in the system will improve in parallel with technical enhancements.

Limitations

Our study had several limitations. First, the sample size was small, with only eight emergency physicians, owing to the current shortage of physicians in South Korea following the temporal mass resignation of residents[19,20]. Therefore, further studies with larger sample sizes are warranted. Second, the study was conducted at a single institution, which may limit the generalizability of our findings. Third, the reliance on self-reported measures introduces subjectivity. Fourth, we developed a measure of concerns regarding LLMs based on the eight concerns identified by Spotnitz[4], with each concern being represented by a single survey item. However, this approach enabled us to directly capture specific physician concerns; the use of single-item measures precluded the assessment of psychometric properties, such as validity and reliability. Fifth, although the EHR logs enabled the calculation of actual documentation time in the postimplementation phase, no reliable start time marker was recorded in the preimplementation phase, precluding a direct comparison between the actual and perceived workload across both periods. Finally, the documentation time was assessed using recall-based methods[21].

Conclusion

In summary, this study implemented an on-premise LLM system for drafting ED discharge notes in a clinical setting and evaluated physician acceptance by comparing perceptions before and after the deployment of the system. The survey results revealed strong physician acceptance, as both the initial concerns about using the LLM in clinical documentation and the perceived workload of writing a discharge note subsided within days, highlighting the practical value of the system. Open-ended feedback highlighted three priorities: improving factual accuracy, expanding reference/EHR inputs, and better tailoring complex or atypical cases. These insights will guide the next Y-KNOT iteration, integrating richer clinical data and upgrading to more capable LLMs, to further reduce the workload and strengthen user trust in future evaluations.

Methods

Participants and data sources

We recruited 15 attending ED physicians at Severance Hospital, Seoul, South Korea, and eight of them voluntarily participated in this study. We obtained online informed consent and conducted a survey prior to the implementation of the LLM system. They received instructions on the Y-KNOT-EDN system and were required to use it to draft discharge notes throughout the study period. Two follow-up surveys were also conducted.

The survey was developed by two researchers (S.L. and J.W.S.) using the SurveyMonkey online survey platform (San Mateo, California, USA; http://www.surveymonkey.com). The study was approved by the Institutional Review Board of Severance Hospital, Yonsei University Health System (No. 4-2024-1622), and was conducted in accordance with the Declaration of Helsinki and all relevant institutional and national guidelines and regulations.

System overview

Y-KNOT-EDN is powered by Llama-3-8B[22], which was fine-tuned on 90 GB of medical and 9 GB of general Korean–English corpora from nonpatient sources and instruction-tuned with real-world clinical records. The model is fully integrated with the hospital's EHR through fast healthcare interoperability, resource-based data exchange, and predefined application programming interfaces. ED discharge draft generation is triggered in real time by clicking an "AI note" button embedded within the order communication system accessed on physicians' desktops. The physicians subsequently reviewed and verified the drafts before the final confirmation. All processing occurred within the hospital's secure on-premise environment with no external transmission of patient data, thus meeting South Korea's strict data sovereignty requirements. The detailed system architecture, model training process, and evaluation results are described in our previous technical report[12].

Survey phases

To conduct a longitudinal evaluation of the clinical application of the Y-KNOT-EDN, we conducted a three-phase survey-based study (Fig. 1).

- *T1 (Preimplementation)*: Participants provided baseline data that included their concerns about using LLMs in clinical documentation and their perceived workload associated with manually drafting ED discharge notes. The measures of concerns related to the use of LLMs in clinical documentation were adapted from Spotnitz et al.[4] Spotnitz et al. surveyed 30 physicians using open-ended questions regarding their concerns regarding LLMs in healthcare and identified eight key dimensions: false information, worsening patient care, data bias, loss of human control, impersonal drafts, legal/ethical issues, privacy, and worsening clinicians' reasoning. We used these dimensions as measurement items, which were assessed on a 5-point Likert scale. The perceived workload of ED discharge documentation was assessed using the NASA-TLX[23], a widely adopted instrument that evaluates workload across six dimensions: mental demand, physical demand, temporal demand, dissatisfaction with one's performance, effort required for accomplishment, and frustration. Each dimension was rated on a scale ranging from 0 to 20. In addition, demographic information was collected, including sex, age, work experience in the ED, and the average time required for manual documentation.
- *T2 (3 d postimplementation)*: At this point, concerns and workload were reassessed along with system acceptance measures—perceived usefulness[24], attitude[25], intention to use[26], and intention to delegate drafting[27]—by adapting measurement items from prior studies.

Variables (phase)	Dimensions	Items
Concerns about using LLM in clinical settings* (T1, T2, T3)	False information	I worry that using the LLM system for discharge note documentation may generate inaccurate or false information.
	Worsening patient care	I worry that using the LLM system for discharge note documentation may adversely affect patient care.
	Data bias	I worry that using the LLM system for discharge note documentation may distort content due to low-quality and biased training data.
	Loss of control	I worry that using the LLM system for discharge note documentation may make it difficult for physicians to control and supervise the process.
	Impersonal draft	I worry that using the LLM system for discharge note documentation may produce impersonal records with low levels of empathy.
	Legal/ethical issue	I worry that using the LLM system for discharge note documentation may lead to ethical and legal issues.
	Privacy	I worry that using the LLM system for discharge note documentation may raise privacy concerns.
	Worsening physicians' reasoning	I worry that using the LLM system for discharge note documentation may reduce physicians' opportunities to interpret and synthesize data.
Perceived workload† (T1, T2, T3)	Mental demand	How mentally demanding was the discharge note documentation task?
	Physical demand	How physically demanding was the discharge note documentation task?
	Temporal demand	How hurried or rushed was the pace of the discharge note documentation task?
	Dissatisfaction with performance‡	How successful were you in accomplishing the discharge note documentation?
	Effort for accomplishment	How hard did you have to work to accomplish your level of performance in the discharge note documentation?
	Frustration	How insecure, discouraged, irritated, stressed, and annoyed did you feel during the discharge note documentation task?
Perceived usefulness* (T2, T3)	–	The LLM system increases productivity in discharge note documentation.
		The LLM system is effective for discharge note documentation.
		I believe that the LLM system is useful for discharge note documentation.
Attitude* (T2, T3)	–	Using the LLM system for discharge note documentation is a good idea.
		Using the LLM system for discharge note documentation is a wise idea.
		I like using the LLM system for discharge note documentation.
		Using the LLM system for discharge note documentation makes me feel good.
Intention to use* (T2, T3)	–	I intend to use the LLM system for discharge note documentation.
		I have an intention to use the LLM system for discharge note documentation.
		I will try to use the LLM system for discharge note documentation.
Intention to delegate* (T2, T3)	–	I plan to delegate discharge note drafting to the LLM system.
		I intend to delegate discharge note drafting to the LLM system.
		I have chosen to use the LLM system for discharge note drafting.
Documentation time (T1, T3)	–	How much time does it take to write a single discharge note?
Experience of shutdown (T3)	–	During your work period, did you experience any downtime or errors that rendered Y-KNOT-EDN unusable? (Yes/No)
		(If "Yes") How challenging was the task during the shutdown?

Table 1. Measurements. *These items are measured using a 1–5 Likert scale. †Workload is measured on a 0–20 scale. The actual question was phrased positively; its score was calculated by subtracting the reported value from a maximum score of 20.

- *T3 (5 weeks postimplementation)*: This phase repeated the survey items from T2 with the addition of questions related to the documentation time per note when using the system. Initially, the third survey was planned to be administered four weeks after the implementation of the Y-KNOT-EDN in the clinical workflow. However, an unexpected one-week system shutdown occurred one week after implementation, compelling ED physicians to revert to manual documentation. Consequently, the survey was postponed to 5 weeks postlaunch, and additional questions were included to capture the impact of this shutdown. Specifically, participants were asked whether they experienced a temporary shutdown of the system during their work period. Those who had experienced the shutdown were subsequently asked to rate the difficulty of manually drafting discharge notes during downtime on a 5-point scale, with 1 indicating "not at all difficult" and 5 indicating "very difficult." Open-ended feedback was also collected through four verbatim questions designed to explore users' subjective experiences with the system: (1) "What aspects of using Y-KNOT-EDN were most helpful to you? (e.g., reduced writing time, decreased workload, system convenience, etc.)" (2) "What were your impressions when reviewing the discharge note drafts generated by Y-KNOT-EDN? (e.g., accuracy, appropriateness of expressions, need for revision, etc.)" (3) "In your opinion, how did Y-KNOT-EDN affect physicians' work efficiency and patient care processes?" (4) "What additional features or improvements would you like to see in future versions of Y-KNOT-EDN to enhance its effectiveness?"

The full survey questions and the scoring scales are summarized in Table 1.

Hypotheses

Based on this study design, we formulated three detailed hypotheses: (1) concerns regarding the use of LLMs in clinical documentation, which physicians experienced at T1 would be alleviated at T2 or T3; (2) the use of the Y-KNOT-EDN would lead to a decrease in the perceived workload of drafting discharge notes when comparing T1 with T2 or T3; and (3) the four acceptance measures of the Y-KNOT-EDN would remain high (4-to-5) from T2 to T3.

Statistical analysis

The statistical analysis employed a Friedman test to assess changes across the three time points (T1, T2, and T3) for both the individual LLM concern items (the overall concern score) and the six workload dimensions (their aggregated overall scores). When the Friedman test indicated a significant effect, post hoc Wilcoxon signed-rank tests with Holm–Bonferroni correction were conducted to perform pairwise comparisons between time points. In addition, a paired t-test was used to compare the time required to complete a note between T1 and T3 and the perceived usefulness, attitude, intention to use, and intention to delegate drafting between T2 and T3. All statistical analyses and visualizations were conducted using Python version 3.12 (Python Software Foundation, Delaware, USA).

Data availability

The raw data are disclosed in Supplementary Table S5 and S6.

Received: 17 June 2025; Accepted: 14 October 2025
Published online: 24 November 2025

References

1. Boussina, A. et al. Large Language models for more efficient reporting of hospital quality measures. *NEJM AI*. **1**, AIcs2400420 (2024).
2. Ong, J. C. L. et al. Medical ethics of large Language models in medicine. *NEJM AI*, AIra2400038 (2024).
3. Gallifant, J. et al. The TRIPOD-LLM reporting guideline for studies using large Language models. *Na. Med.* 1–10 (2025).
4. Spotnitz, M. et al. A survey of clinicians' views of the utility of large Language models. *Appl. Clin. Inf.* **15**, 306–312. https://doi.org/10.1055/a-2281-7092 (2024).
5. Tripathi, S., Sukumaran, R. & Cook, T. S. Efficient healthcare with large Language models: optimizing clinical workflow and enhancing patient care. *J. Am. Med. Inform. Assoc.* **31**, 1436–1440 (2024).
6. Hartman, V. et al. Developing and evaluating large Language model–generated emergency medicine handoff notes. *JAMA Netw. Open.* **7**, e2448723–e2448723 (2024).
7. Hartman, V. C. et al. A method to automate the discharge summary hospital course for neurology patients. *J. Am. Med. Inform. Assoc.* **30**, 1995–2003 (2023).
8. Chua, C. E. et al. Integration of customised LLM for discharge summary generation in real-world clinical settings: a pilot study on RUSSELL GPT. *The Lancet Reg. Health–Western Pacific* **51** (2024).
9. Heilmeyer, F. et al. Viability of open large Language models for clinical Documentation in German health care: Real-World model evaluation study. *JMIR Med. Inf.* **12**, e59617 (2024).
10. Association, A. M. AMA Augmented Intelligence Research: Physician Sentiments Around the Use of AI in Health Care: Motivations, Opportunities, Risks, and Use Cases – Shifts from 2023 to 2024. (2025).
11. Park, S. H. & Langlotz, C. P. Crucial role of Understanding in Human-Artificial intelligence interaction for successful clinical adoption. *Korean J. Radiol.* **26**, 287–290 (2025).
12. Kim, H. et al. A Bilingual On-premise AI agent for Clinical Drafting: Seamless EHR integration in the Y-KNOT Project. medRxiv:2025.2004. 25325003 (2025). (2003).
13. Lambert, S. I. et al. An integrative review on the acceptance of artificial intelligence among healthcare professionals in hospitals. *NPJ Digit. Med.* **6**, 111 (2023).
14. Jindal, J. A., Lungren, M. P. & Shah, N. H. Ensuring useful adoption of generative artificial intelligence in healthcare. *J. Am. Med. Inform. Assoc.* **31**, 1441–1444. https://doi.org/10.1093/jamia/ocae043 (2024).
15. Gandhi, T. K. et al. How can artificial intelligence decrease cognitive and work burden for front line practitioners? *JAMIA open* **6**, ooad079 (2023).
16. Tierney, A. A. et al. Ambient artificial intelligence scribes to alleviate the burden of clinical Documentation. *NEJM Catalyst Innovations Care Delivery* **5**, (2024). CAT. 23.0404.
17. Sellergren, A. et al. Medgemma technical report. *arXiv preprint arXiv:2507.05201* (2025).
18. OpenAI *Introducing gpt-oss*. https://openai.com/ko-KR/index/introducing-gpt-oss/ (2025).
19. McCurry, J. South Korean Doctors threaten mass resignation. *Lancet* **403**, 1124 (2024).
20. Moon, J. & Lee, J. Y. Why I decide to leave South Korea healthcare system. *The Lancet Reg. Health–Western Pacific* **52** (2024).
21. Murad, M. H. et al. Measuring Documentation burden in healthcare. *J. Gen. Intern. Med.* 1–12 (2024).
22. Meta *Introducing Meta Llama 3: The most capable openly available LLM to date*, (2024). https://ai.meta.com/blog/meta-llama-3/.
23. Hart, S. G. in *Proceedings of the human factors and ergonomics society annual meeting.* 904–908 (Sage publications Sage CA: Los Angeles, CA).
24. Venkatesh, V. & Davis, F. D. A theoretical extension of the technology acceptance model: four longitudinal field studies. *Manage. Sci.* **46**, 186–204 (2000).
25. Davis, F. D. & Perceived Usefulness Perceived ease of Use, and user acceptance of information technology. *MIS Q.* **13**, 319–340. https://doi.org/10.2307/249008 (1989).
26. Ajzen, I., Fishbein, M., Lohmann, S. & Albarracín, D. The influence of attitudes on behavior. *The handbook of attitudes, volume 1: Basic principles*, 197–255 (2018).
27. Stout, N., Dennis, A. R. & Wells, T. M. The Buck stops there: the impact of perceived accountability and control on the intention to delegate to software agents. *AIS Trans. Hum Comput Interact.* **6**, 1–15 (2014).

Acknowledgements

This research was supported by a grant of the MD-Phd/Medical Scientist Training Program through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea. This research was also supported by a grant of the Korea Health Technology R&D Project through the

KHIDI (grant number: RS-2023-KH135326).

Author contributions

S.L. developed the study protocol, analysed data and edited manuscript.J.W.S. analysed data, wrote and edited manuscript. S.C.Y. developed the study protocol and edited manuscript.J.H.K. gathered the interviewees and edited manuscriptS.L. and J.W.S. are co-first authors and have contributed equally to this work. S.C.Y. and J.H.K. are corresponding authors and have contributed equally to this work. The corresponding authors attest that all listed authors meet the authorship criteria and that others who met the criteria have not been omitted.

Funding

This research was supported by a grant from the MD-PhD/Medical Scientist Training Program through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health and Welfare, Republic of Korea. This research was supported by a grant from the Korea Health Technology R&D Project through the KHIDI (grant number: RS-2023-KH135326).

Declarations

Competing interests

S.C.Y reports grants from Daiichi Sankyo. He is a coinventor of granted Korea Patent DP-2023-1223 and DP-2023-0920, and pending Patent Applications DP-2024-0909, DP-2024-0908, DP-2022-1658, DP-2022-1478, and DP-2022-1365 unrelated to current work. S.C.Y. is a chief executive officer of PHI Digital Healthcare. Other authors have no potential conflicts of interest to disclose.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-24659-4.

Correspondence and requests for materials should be addressed to S.C.Y. or J.H.K.