

Original Paper

Comparison of the Quality of Discharge Letters Written by Large Language Models and Junior Clinicians: Single-Blinded Study

Joshua Yi Min Tung^{1,2}, MPH; Sunil Ravinder Gill¹, MBBS; Gerald Gui Ren Sng^{2,3}, MMed; Daniel Yan Zheng Lim^{2,4}, MTech; Yuhe Ke^{2,5}, MMed; Ting Fang Tan⁶, MBBS; Liyuan Jin⁷, BEng; Kabilan Elangovan⁸, BEng; Jasmine Chiat Ling Ong⁹, PharmD; Hairil Rizal Abdullah^{2,5}, PhD; Daniel Shu Wei Ting^{6,7,8}, PhD; Tsung Wen Chong¹, DPhil

¹Department of Urology, Singapore General Hospital, Singapore, Singapore

²Data Science and Artificial Intelligence Laboratory, Singapore General Hospital, Singapore, Singapore

³Department of Endocrinology, Singapore General Hospital, Singapore, Singapore

⁴Department of Gastroenterology and Hepatology, Singapore General Hospital, Singapore, Singapore

⁵Department of Anesthesiology, Singapore General Hospital, Singapore, Singapore

⁶Singapore Eye Research Institute, Singapore National Eye Center, Singapore, Singapore

⁷Duke-NUS Medical School, Singapore, Singapore

⁸Artificial Intelligence Office, Singapore Health Services, Singapore, Singapore

⁹Division of Pharmacy, Singapore General Hospital, Singapore, Singapore

Corresponding Author:

Joshua Yi Min Tung, MPH

Department of Urology

Singapore General Hospital

16 College Road, Block 4 Level 1

Singapore, 169854

Singapore

Phone: 65 62223322

Email: joshua.tung@gmail.com

Abstract

Background: Discharge letters are a critical component in the continuity of care between specialists and primary care providers. However, these letters are time-consuming to write, underprioritized in comparison to direct clinical care, and are often tasked to junior doctors. Prior studies assessing the quality of discharge summaries written for inpatient hospital admissions show inadequacies in many domains. Large language models such as GPT have the ability to summarize large volumes of unstructured free text such as electronic medical records and have the potential to automate such tasks, providing time savings and consistency in quality.

Objective: The aim of this study was to assess the performance of GPT-4 in generating discharge letters written from urology specialist outpatient clinics to primary care providers and to compare their quality against letters written by junior clinicians.

Methods: Fictional electronic records were written by physicians simulating 5 common urology outpatient cases with long-term follow-up. Records comprised simulated consultation notes, referral letters and replies, and relevant discharge summaries from inpatient admissions. GPT-4 was tasked to write discharge letters for these cases with a specified target audience of primary care providers who would be continuing the patient's care. Prompts were written for safety, content, and style. Concurrently, junior clinicians were provided with the same case records and instructional prompts. GPT-4 output was assessed for instances of hallucination. A blinded panel of primary care physicians then evaluated the letters using a standardized questionnaire tool.

Results: GPT-4 outperformed human counterparts in information provision (mean 4.32, SD 0.95 vs 3.70, SD 1.27; $P=.03$) and had no instances of hallucination. There were no statistically significant differences in the mean clarity (4.16, SD 0.95 vs 3.68, SD 1.24; $P=.12$), collegiality (4.36, SD 1.00 vs 3.84, SD 1.22; $P=.05$), conciseness (3.60, SD 1.12 vs 3.64, SD 1.27; $P=.71$), follow-up recommendations (4.16, SD 1.03 vs 3.72, SD 1.13; $P=.08$), and overall satisfaction (3.96, SD 1.14 vs 3.62, SD 1.34; $P=.36$) between the letters generated by GPT-4 and humans, respectively.

Conclusions: Discharge letters written by GPT-4 had equivalent quality to those written by junior clinicians, without any hallucinations. This study provides a proof of concept that large language models can be useful and safe tools in clinical documentation.

(*J Med Internet Res* 2024;26:e57721) doi: [10.2196/57721](https://doi.org/10.2196/57721)

KEYWORDS

artificial intelligence; AI; discharge summaries; continuity of care; large language model; LLM; junior clinician; letter writing; single-blinded; ChatGPT; urology; primary care; fictional electronic record; consultation note; referral letter; simulated environment

Introduction

Large language models (LLMs) are artificial intelligence (AI) constructs with the capability to parse and generate human-like text based on pretraining with a vast corpus of unstructured text data. LLMs have shown promise in parsing large corpi of text data and generating human-like text across various domains, including health care [1].

One of the most promising applications of LLMs in health care is for automating tasks traditionally performed by health care professionals, such as writing discharge summaries and letters [2]. Despite the fact that such documents serve as a critical communication tool between hospital specialists and primary care providers (PCPs) in ensuring continuity of care, they are time-consuming to write, are often tasked to junior doctors, and are underprioritized when balanced against direct clinical care [3,4].

Prior studies have shown that inpatient hospital discharge summary quality is inadequate in many of the domains/elements required by The Joint Commission and endorsed at the Transitions of Care Consensus Conference (TOCCC) [3]. Many elements, including the provision of follow-up care plans, are of particular importance to patients being discharged to primary care. Improving discharge summary timeliness was shown to have positive effects on the quality of content and the transmission rate to PCPs [5,6].

The automation of inpatient discharge summaries has previously been explored using electronic discharge summary (EDS) programs. Maslove et al [7] showed no differences in PCP-reported overall quality between summaries generated by an EDS program and conventional dictated summaries. However, these studies assessed communication between inpatient and outpatient providers for single episodes of hospital admissions rather than summative information regarding a patient's clinical condition across a longer follow-up, as would be the case with specialist outpatient care.

Moreover, the applications of LLMs in outpatient clinical documentation remain largely unexplored. Unlike inpatient admissions, which are typically self-contained and time-limited, outpatient specialist care can have more complex clinical

documentation with longitudinal health records that may span across years. It remains to be seen if LLMs are able to parse clinical information with a strong temporal dimension and produce high-quality output.

There is currently no consensus on the single best way to objectively evaluate clinically relevant LLM output. However, approaches comparing LLM output against that generated by clinicians can help with assessment of a benchmark of "human-like" or "clinician-like" performance, which have been described in the related literature [8]. In addition, independent expert evaluation of LLM output has also been used for physician evaluation of LLM output [9]. We sought to combine both approaches in this study.

The aims of this study were to (1) evaluate the feasibility and quality of an LLM in writing discharge letters for patients discharged from a specialist outpatient clinic and (2) compare the discharge letters generated against similar letters written by human junior clinicians.

Methods

Ethical Considerations

This study was conducted in a simulated environment using only fictional patient data. As the use of fictional data does not fall under local Human Biomedical Research Act regulations [10], ethical approval was not required.

Data Generation

Fictional patient data were generated by clinicians from the Singapore General Hospital's Department of Urology, mimicking typical electronic medical records (EMRs) of 5 common patient groups seen at the outpatient clinic in a tertiary center. The time horizon for follow-up ranged from 6 months to 6 years. Data included initial referral letters from PCPs or emergency departments, initial and follow-up urology consultation notes, referrals to other departments, correspondence notes, and discharge summaries from relevant admissions. Information on the fictional cases is presented in [Table 1](#). Each case varied in complexity, ranging between 4 and 10 documents, and included 1238-3009 characters (424-1110 tokens).

Table 1. Fictional case-mix data used for assessment.

| Simulated case | Initial referral | Clinical condition | Follow-up duration | Additional/incidental clinical information |
|----------------|------------------------------|------------------------------|--------------------|---|
| 1 | Elevated PSA ^a | Benign prostatic hyperplasia | 6 years | Initially declined biopsy and opted for PSA monitoring, but ultimately underwent a transperineal prostate biopsy and developed acute retention of urine postoperatively |
| 2 | LUTS ^b | Overactive bladder | 3 years | Failed behavioral modification measures and experienced adverse effects from multiple lines of anticholinergic medications |
| 3 | Asymptomatic micro-hematuria | Renal calculus | 1.5 years | Incidental pancreatic lesion on imaging and was referred to general surgery for further evaluation |
| 4 | Ureteric colic | Distal ureteric stone | 6 months | Failed medical expulsive therapy and underwent ureteroscopy and lithotripsy; also noted to have hypertension perioperatively |
| 5 | Erectile dysfunction | Erectile dysfunction | 2.5 years | Noted to have abnormal ECG ^c and was referred to cardiology for further assessment prior to initiation of PDE-5 ^d inhibitors |

^aPSA: prostate-specific antigen.

^bLUTS: lower urinary tract symptoms.

^cECG: electrocardiography.

^dPDE-5: phosphodiesterase-5.

Discharge Letter Generation

GPT-4 is a state-of-the-art LLM developed by OpenAI with a parameter count of 1.76 trillion, and has exhibited human-level performance on professional and academic benchmarks [11]. GPT-4, with an 8000-token context length, was provided with the fictional EMRs and tasked to write discharge letters to PCPs. The prompt instructed GPT-4 to assume the role of a physician assistant in a urology clinic, specifying the context of the task (“this patient is being discharged”) and the target audience (“meant to be read by the general practitioner”). To reduce the risk of hallucination, prompts for safety included specific instructions to include only information provided in the fictional EMR [12]. Desired content guidelines were provided to standardize the structure of the discharge letter. Prompts for style included instructions to write in prose, in a cordial and concise manner, and not to exceed half a page of text unless necessary. To evaluate the generative capacity of the LLM, only the above instructions (but no examples) were provided to GPT-4, which is a technique known as “zero-shot” prompting [13].

Three separate discharge letters were generated for each fictional case. One was generated by GPT-4 and the comparators were two letters that were written concurrently by junior clinicians from the Department of Urology using the same set of fictional patients. For the comparator letters, use of generative AI tools or other automated summarization methods was disallowed.

Full prompt instructions are shown in [Multimedia Appendix 1](#). Sample medical records and discharge letters are shown in [Multimedia Appendix 2](#).

Comparison of Discharge Letters

All letters underwent an initial independent screening by two study team members (JYMT and SRG) for factual inaccuracy and grammatical errors. For letters generated by GPT-4, this included screening for instances of hallucination by the language model. For cases 3 and 4, letters were also assessed with respect to their attention to secondary problems beyond the prompted surgical issue based on the presence or absence of follow-up recommendations for other medical issues found during the course of the patient’s treatment.

A panel of 5 senior primary care physicians was presented with the discharge letters written by GPT-4 and junior clinicians in a blinded fashion. The participating physicians on the expert panel had an average clinical practice experience of 34.4 years. They compared the letters produced by GPT-4 and junior clinicians using a standardized rubric, but they were not informed that one of the letters was LLM-generated. Letters were evaluated on a 5-point Likert scale for completeness of information, conciseness, clarity, collegiality, whether follow-up care plans were articulated, and a single overall satisfaction question ([Figure 1](#)). Since there are no validated tools for assessment of outpatient discharge letter quality, questionnaire items were selected based on elements endorsed by the TOCCC for inpatient summaries.

Figure 1. Scoring questionnaire.

| Question | Strongly Agree | Agree | Neutral | Disagree | Strongly Disagree |
|--|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| 1. Information about the patient's diagnosis and clinical course is satisfactory. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 2. The discharge summary is appropriately concise. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 3. Information in the discharge summary is presented with clarity. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 4. The discharge summary is collegial. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 5. Recommendations for follow-up are clear. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 6. Overall, I would be happy to receive this discharge summary as the patient's next managing physician. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

Data Analysis

SPSS version 26.0 (IBM Corp) was used for the statistical analysis of quantitative data. Total scores and arithmetic means with SDs were calculated for all responses as a whole and for responses to each individual item. As the data were nonnormally distributed (Shapiro-Wilk test $P < .001$), differences in means were compared using the Mann-Whitney U test.

Results

Discharged letters generated by GPT-4 averaged 276 words (1462 characters), in comparison to 155 words (737 characters) for letters written by junior clinicians. No instances of

hallucination and no grammatical errors were present in the letters generated by GPT-4. While no factual inaccuracies were found in the letters written by the junior clinicians, they made an average of 1.7 grammatical errors per letter.

Each panel member scored all 15 letters, for a total of 75 responses. The mean total score per letter was 22.99 (SD 6.26). GPT-4 scored higher on question items 1 (information), 3 (clarity), 4 (collegiality), 5 (follow-up plans), and 6 (overall satisfaction), and received higher total mean scores as compared to those of human subjects, although this difference was only statistically significant for item 1. GPT-4 received lower scores on item 2 (conciseness). The complete results are shown in [Table 2](#).

Table 2. Comparison of mean scores for each evaluation item of discharge letters written by GPT-4 and junior clinicians.

| Item | GPT-4 (n=25), mean (SD) | Junior clinicians (n=50), mean (SD) | P value |
|----------------------|-------------------------|-------------------------------------|---------|
| Information | 4.32 (0.95) | 3.70 (1.27) | .03 |
| Conciseness | 3.60 (1.12) | 3.64 (1.27) | .71 |
| Clarity | 4.16 (0.94) | 3.68 (1.24) | .12 |
| Collegiality | 4.36 (1.00) | 3.84 (1.22) | .05 |
| Recommendations | 4.16 (1.03) | 3.72 (1.13) | .08 |
| Overall satisfaction | 3.96 (1.14) | 3.62 (1.34) | .36 |
| Total | 24.56 (5.13) | 22.20 (6.66) | .20 |

GPT-4 also had some unanticipated strengths when compared against the letters written by the junior clinicians. In particular, GPT-4 demonstrated attention to secondary problems that were not explicitly included in the original prompt. In cases 3 and 4, GPT-4 provided follow-up recommendations based on the given fictional records for the incidental pancreatic lesion (“we recommend annual imaging studies to monitor the renal stone and the pancreatic lesion, in collaboration with the General Surgery team”) and perioperative hypertension (“he is

recommended to continue with Amlodipine 5mg daily for his hypertension”). In contrast, although the junior clinicians also made mention of the pancreatic lesion and subsequent general surgery referral, they did not provide follow-up recommendations to the PCP.

Discussion

Principal Findings

To our knowledge, this study represents the first blinded comparison of discharge letters written by an advanced LLM versus human physicians.

Our analysis showed that GPT-4 performed comparably against junior clinicians from an inpatient specialist department, with equivalent scores in clarity, collegiality, and ability to articulate follow-up recommendations. GPT-4 also demonstrated the ability to generate a holistic patient summary with attention to secondary medical issues and made fewer grammatical errors than human comparators. The higher score in information provision highlights the capability of language models with strong embeddings to capture and present important details from patient EMRs. However, this appears to have come at the expense of a relative lack of conciseness, as indicated by GPT-4 receiving a lower score in this domain. More specific metaprompting with instructions for word limits could overcome this limitation. Nevertheless, the importance of an LLM's ability to rapidly digest and summarize a large free-text corpus cannot be overstated, which offers the potential for more consistent information relay between health care providers.

These findings echo prior work in the field comparing EDSs against traditional dictated summaries. While earlier digitization efforts were limited by technological capabilities of the era, advances in natural language processing and linguistic ability allow modern language models to write fluently and mimic human prose convincingly. While the potential of language models in this area of health care has been discussed [2], these first results prove that LLMs do have the capability to automate time-consuming tasks for health care professionals. This is a transformative opportunity that could be a useful tool in bridging communications between PCPs and inpatient specialists by ensuring the quality and accuracy of discharge letters. Clinicians burdened with documentation in addition to clinical duties stand to benefit immensely from automation of these tasks, allowing more time for patient contact and clinical care. Likewise, patients could benefit from consistent, timely handovers and a more robust system of continuity of care, instead of the vagaries of the junior clinicians writing their discharge letters. With appropriate refinement of prompt instructions, we postulate that this transformative benefit can extend beyond automating the documentation work of physicians to other health care professionals (eg, nurses and pharmacists).

Data Availability

The data sets generated and/or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

JYMT, GGRS, DYZL, and TWC contributed to the conceptualization of the study. JYMT performed the data curation and formal analysis. SRG contributed to the writing of the original draft. HRA, DSWT, and TWC supervised the study. YK, TFT, LJ, KE, and JCLO contributed to revisions and editing of the manuscript.

Limitations and Prospects

The exclusive reliance on fictional patient data in this study, while necessary for ethical and patient data considerations, might not capture the real-world complexities in patient EMRs. We attempted to overcome this limitation by having the fictional EMRs written by physicians who also regularly see patients in the urology outpatient clinic on a day-to-day basis and enter clinical notes into actual patient EMRs. The small number of panel assessors as well as the use of clinical cases from a single surgical specialty may reduce the generalizability of these findings. No evaluation was performed as to whether the panel assessors inferred that any of the letters were LLM-generated. The time taken for junior clinicians to write discharge letters was not measured, precluding comparison in terms of time savings against an LLM. In addition, despite the advantages of LLM tools, misinformation phenomena such as hallucination and concept drift remain areas of concern. We propose that such an AI tool should be used to augment, not replace, human-written discharge letters, and that physician oversight should still be required before discharge documents are handed to patients. Users of LLMs as productivity aids in clinical medicine must also be aware of prevailing privacy protection policies, and closed-access LLM implementations (eg, on a private server) may be needed for regulatory compliance. While automating discharge letter writing may undermine the clinical reasoning process that physicians undertake [14], we believe that LLMs implemented with a "human in the loop" can eventually exert a long-term training effect, reinforcing these cognitive skill sets [15].

Future research using actual patient EMRs will be necessary to confirm these preliminary findings. Studies may be expanded to care across different medical specialties to ascertain the ability of language models to comprehend abbreviations and concepts from different disciplines. Fine-tuning a language model on medical information and patient EMRs and refining metaprompts are avenues to further improve the quality of discharge letters generated by LLMs and to balance conciseness with information density.

Conclusion

The emergence of AI in health care promises a paradigm shift in the way clinical medicine is practiced. Our study provides insight into AI's capacity to optimize the continuity of patient care. As evaluation of language models for clinical applications continues to advance, we propose benchmarking their performance against human counterparts to determine feasibility and assess output quality.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Prompt instructions.

[\[PNG File , 117 KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

Sample fictional clinical records and discharge letters.

[\[DOCX File , 56 KB-Multimedia Appendix 2\]](#)

References

1. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. *Nat Med*. Aug 17, 2023;29(8):1930-1940. [doi: [10.1038/s41591-023-02448-8](https://doi.org/10.1038/s41591-023-02448-8)] [Medline: [37460753](https://pubmed.ncbi.nlm.nih.gov/37460753/)]
2. Patel SB, Lam K. ChatGPT: the future of discharge summaries? *Lancet Digit Health*. Mar 2023;5(3):e107-e108. [FREE Full text] [doi: [10.1016/S2589-7500\(23\)00021-3](https://doi.org/10.1016/S2589-7500(23)00021-3)] [Medline: [36754724](https://pubmed.ncbi.nlm.nih.gov/36754724/)]
3. Horwitz LI, Jenq GY, Brewster UC, Chen C, Kanade S, Van Ness PH, et al. Comprehensive quality of discharge summaries at an academic medical center. *J Hosp Med*. Aug 22, 2013;8(8):436-443. [FREE Full text] [doi: [10.1002/jhm.2021](https://doi.org/10.1002/jhm.2021)] [Medline: [23526813](https://pubmed.ncbi.nlm.nih.gov/23526813/)]
4. Rao P, Andrei A, Fried A, Gonzalez D, Shine D. Assessing quality and efficiency of discharge summaries. *Am J Med Qual*. Sep 03, 2005;20(6):337-343. [doi: [10.1177/1062860605281078](https://doi.org/10.1177/1062860605281078)] [Medline: [16280397](https://pubmed.ncbi.nlm.nih.gov/16280397/)]
5. Weetman K, Spencer R, Dale J, Scott E, Schnurr S. What makes a "successful" or "unsuccessful" discharge letter? Hospital clinician and general practitioner assessments of the quality of discharge letters. *BMC Health Serv Res*. Apr 15, 2021;21(1):349. [FREE Full text] [doi: [10.1186/s12913-021-06345-z](https://doi.org/10.1186/s12913-021-06345-z)] [Medline: [33858383](https://pubmed.ncbi.nlm.nih.gov/33858383/)]
6. Cresswell A, Hart M, Suchanek O, Young T, Leaver L, Hibbs S. Mind the gap: improving discharge communication between secondary and primary care. *BMJ Qual Improv Rep*. Sep 02, 2015;4(1):u207936.w3197. [FREE Full text] [doi: [10.1136/bmjquality.u207936.w3197](https://doi.org/10.1136/bmjquality.u207936.w3197)] [Medline: [26734391](https://pubmed.ncbi.nlm.nih.gov/26734391/)]
7. Maslove DM, Leiter RE, Griesman J, Arnott C, Mourad O, Chow C, et al. Electronic versus dictated hospital discharge summaries: a randomized controlled trial. *J Gen Intern Med*. Sep 16, 2009;24(9):995-1001. [FREE Full text] [doi: [10.1007/s11606-009-1053-2](https://doi.org/10.1007/s11606-009-1053-2)] [Medline: [19609623](https://pubmed.ncbi.nlm.nih.gov/19609623/)]
8. Bernstein IA, Zhang YV, Govil D, Majid I, Chang RT, Sun Y, et al. Comparison of ophthalmologist and large language model chatbot responses to online patient eye care questions. *JAMA Netw Open*. Aug 01, 2023;6(8):e2330320. [FREE Full text] [doi: [10.1001/jamanetworkopen.2023.30320](https://doi.org/10.1001/jamanetworkopen.2023.30320)] [Medline: [37606922](https://pubmed.ncbi.nlm.nih.gov/37606922/)]
9. Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, Chung HW, et al. Large language models encode clinical knowledge. *Nature*. Aug 2023;620(7972):172-180. [FREE Full text] [doi: [10.1038/s41586-023-06291-2](https://doi.org/10.1038/s41586-023-06291-2)] [Medline: [37438534](https://pubmed.ncbi.nlm.nih.gov/37438534/)]
10. Human Biomedical Research Act. Ministry of Health Singapore. URL: <https://www.moh.gov.sg/policies-and-legislation/human-biomedical-research-act> [accessed 2024-07-19]
11. OpenAI, Achiam J, Adler S, Agarwal S, Ahmad L, Akkaya I, et al. GPT-4 Technical Report. arXiv Preprint posted online on March 15, 2023. 2024. [FREE Full text]
12. Bender E, Gebru T, McMillan-Major A, Shmitchell S. On the dangers of stochastic parrots: can language models be too big? In: *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. New York, NY. Association for Computing Machinery; 2021:610-623.
13. Zhou C, He J, Ma X, Berg-Kirkpatrick T, Neubig G. Prompt consistency for zero-shot task generalization. arXiv Preprint posted online on April 29, 2022. 2024. [doi: [10.18653/v1/2022.findings-emnlp.192](https://doi.org/10.18653/v1/2022.findings-emnlp.192)]
14. Preiksaitis C, Sinsky CA, Rose C. ChatGPT is not the solution to physicians' documentation burden. *Nat Med*. Jun 2023;29(6):1296-1297. [doi: [10.1038/s41591-023-02341-4](https://doi.org/10.1038/s41591-023-02341-4)] [Medline: [37169865](https://pubmed.ncbi.nlm.nih.gov/37169865/)]
15. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. *NPJ Digit Med*. 2020;3:17. [doi: [10.1038/s41746-020-0221-y](https://doi.org/10.1038/s41746-020-0221-y)] [Medline: [32047862](https://pubmed.ncbi.nlm.nih.gov/32047862/)]

Abbreviations

- AI:** artificial intelligence
EDS: electronic discharge summary
EMR: electronic medical record
LLM: large language model
PCP: primary care provider

TOCCC: Transitions of Care Consensus Conference

Edited by G Eysenbach, T de Azevedo Cardoso; submitted 25.02.24; peer-reviewed by L Yong, G Tofler, Chi-Shin Wu; comments to author 08.04.24; revised version received 18.04.24; accepted 04.07.24; published 24.07.24

Please cite as:

*Tung JYM, Gill SR, Sng GGR, Lim DYZ, Ke Y, Tan TF, Jin L, Elangovan K, Ong JCL, Abdullah HR, Ting DSW, Chong TW
Comparison of the Quality of Discharge Letters Written by Large Language Models and Junior Clinicians: Single-Blinded Study
J Med Internet Res 2024;26:e57721*

URL: <https://www.jmir.org/2024/1/e57721>

doi: [10.2196/57721](https://doi.org/10.2196/57721)

PMID:

©Joshua Yi Min Tung, Sunil Ravinder Gill, Gerald Gui Ren Sng, Daniel Yan Zheng Lim, Yuhe Ke, Ting Fang Tan, Liyuan Jin, Kabilan Elangovan, Jasmine Chiat Ling Ong, Hairil Rizal Abdullah, Daniel Shu Wei Ting, Tsung Wen Chong. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 24.07.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.