npj | digital medicine

Brief communication

Published in partnership with Seoul National University Bundang Hospital

https://doi.org/10.1038/s41746-024-01336-w

The quality and safety of using generative AI to produce patient-centred discharge instructions

Kristian Stanceski[1,2,3], Sharleen Zhong[4], Xumou Zhang[4], Sam Khadra[2], Marguerite Tracy[5,6,7], Linda Koria[8], Sarita Lo[8], Vasi Naganathan[8,9], Jinman Kim[4], Adam G. Dunn[1] ✉ & Julie Ayre[5]

Patient-centred instructions on discharge can improve adherence and outcomes. Using GPT-3.5 to generate patient-centred discharge instructions, we evaluated responses for safety, accuracy and language simplification. When tested on 100 discharge summaries from MIMIC-IV, potentially harmful safety issues attributable to the AI tool were found in 18%, including 6% with hallucinations and 3% with new medications. AI tools can generate patient-centred discharge instructions, but careful implementation is needed to avoid harms.

Hospital discharge summaries, also known as discharge referrals or clinical handovers, communicate key information about a patient related to their hospital admission. Details often include presenting complaints, investigations, diagnoses, treatments received, procedures completed and instructions for continuity of care post discharge such as new medications, and follow-up actions such as appointments, tests, wound care and others[1]. Discharge summaries are designed to be the main communication tool to support the safe transition from hospital to community care. Most information in the discharge summary is intended for a primary care physician, rather than for the patient. Evidence suggests that while primary care physicians are supportive of patients receiving a copy of their discharge summary[2], only one third of discharge summaries contained patient-centred information[3].

Medication errors after discharge are common. More than half of patients have misunderstandings about indication, dose, or frequency of the medications they take after discharge and these issues are more common among patients with low health literacy[4]. A meta-analysis of emergency department discharge summaries found that only 58% of patients could correctly recall their written discharge summary instructions[5]. In a study of 254 older adults in the United States, 22% of participants did not understand how to take their medication and the rate was 48% among participants with low health literacy[6].

Discharge summaries can be designed or augmented to improve patient safety in care transitions. Patient-centred language in discharge instructions has been associated with lower rates of readmission and fewer patient calls to hospital[7]. Health literacy guidelines recommend reducing medical jargon and using everyday language to improve understanding of health information[8–10]. The Universal Medication Schedule (UMS) is a specific format that explains medication dosage and timing in relation to time periods (morning, noon, evening, bedtime)[11]. While evidence of its impact is mixed, its use is associated with improved medication adherence, particularly for older adults with more complex medication regimens[12–15].

Language models can be used in tools for simplifying online health information for patients, though current evidence does not yet provide a clear picture of its value[16–20]. One study examined the use of ChatGPT for simplifying radiology reports into plain language that could be used by patients and healthcare providers[21]. We know of no studies that have examined the use of generative artificial intelligence models with discharge summaries to generate new patient-centred discharge instructions supporting their medication use and ongoing actions and appointments.

In this study, we evaluated the safety, accuracy and language simplification of patient-centred discharge instructions generated by a GPT-based model. To do this, we developed a prompt to generate patient-centred discharge instructions using GPT-3.5-turbo-16k 2023-07-01-preview version (hereafter, GPT-3.5). Three prompt strategies were developed and evaluated to find a prompt that balanced language simplification with the

[1]Biomedical Informatics and Digital Health, Faculty of Medicine and Health, The University of Sydney, Sydney, NSW, Australia. [2]Royal Prince Alfred Hospital, Sydney Local Health District, Sydney, NSW, Australia. [3]School of Clinical Medicine, UNSW Medicine & Health, St George and Sutherland Clinical Campus, Sydney, NSW, Australia. [4]School of Computer Science, Faculty of Engineering, The University of Sydney, Sydney, NSW, Australia. [5]Sydney Health Literacy Lab, Sydney School of Public Health, Faculty of Medicine and Health, The University of Sydney, Sydney, NSW, Australia. [6]General Practice Clinical School, Faculty of Medicine and Health, The University of Sydney, Sydney, NSW, Australia. [7]Drug Health Services, Blacktown and Mount Druitt Hospitals, Western Sydney Local Health District, Sydney, NSW, Australia. [8]Centre for Education and Research on Ageing, Department of Geriatric Medicine, Concord Repatriation Hospital, Sydney, NSW, Australia. [9]Concord Clinical School, Faculty of Medicine and Health, University of Sydney, Sydney, NSW, Australia. ✉e-mail: adam.dunn@sydney.edu.au

Table 1 | Performance of the AI-generated patient discharge instructions

Measure	Result (*N* = 100)
Medications	
all medications were included in the response, %	90
no additional medications were included in the response, %	97
mean percentage of medications correct in the response, % (std)	85% (25%)
percentage of medications that were correctly specified by type, dose, route, frequency and duration, median (IQR)	100% (81–100%)
Follow-up actions	
all actions were included in the response, %	50
no additional actions were included in the response, %	58
mean percentage of actions that were correct in the response, % (std)	78% (26%)
percentage of actions that were correctly specified, median (IQR)	86% (67–100%)

correctness of medications and follow up actions in the AI-generated response (see "Methods"). Discharge summaries were used as reference documents from which to generate responses. Clinicians then compared the descriptions of medications and follow-up actions for 100 pairs of AI-generated discharge instructions and their original discharge summaries.

The median length of the original discharge summaries was 1506 words (interquartile range [IQR] 1096–1987). The median number of medications was 9 (IQR 6 to 12) and the median number of actions was 5 (IQR 3–7). Across the original discharge summaries, the mean grade reading level was 10.7 (Standard deviation [SD] 0.5) and the mean language complexity was 40.3% (SD 3.9). The patients were generally older, where patients over 60 comprised 48% of examples.

The AI-generated responses were shorter and simpler than the original discharge summaries. The median length of the responses was 267 words (IQR 197–355), with a grade reading level of 10.1 (SD 1.0) and an average language complexity of 31.2% (SD 4.4). Grade reading level ($p < 0.001$, $t = 5.96$) and language complexity ($p < 0.001$, $t = 15.7$) were both lower in the patient-centred discharge instructions than in the original discharge summaries. Medications were able to be produced in UMS format for 25% (IQR 0–50%) of medications. While the results show a significant reduction in grade reading level and language complexity, the proportion that could be written in UMS format and were correctly represented in UMS format in the AI-generated response was relatively low, suggesting that future studies in the area may wish to consider additional ways to measure how outputs can be best aligned with patient needs and health literacy levels.

Clinicians including pharmacists and primary care physicians compared the text of the original discharge summaries and the patient-centred discharge instructions. Responses captured most of the relevant medications and follow-up actions correctly (Table 1). For example, the median of correctly summarised medications in the patient instructions from the original discharge summary was 100% (IQR 81–100%), while for follow-up actions was 86% (IQR 67–100%). The responses rarely added medications that were not in the original discharge summary (3% of cases) but introduced new actions in 42% of cases (Supplementary Tables 3, 4).

There were a range of safety issues identified in the responses (Fig. 1). Safety issues attributable to the AI-generated response were identified in 18% (18 of 100) of the patient-centred discharge instructions. Other issues that were considered less severe and unlikely to cause harm were identified in 28% (28 of 100). In one case, an AI-generated response included 'Carbamazepine 400 mg: Take 2 tablets by mouth twice daily', whereas the original discharge summary had 'one 400 mg tablet twice daily' (Supplementary Tables 3, 4). In a post-hoc analysis of factors associated with safety issues, we found no evidence of differences relative to patient age, gender, total medications, or type of care service (Supplementary Table 1).

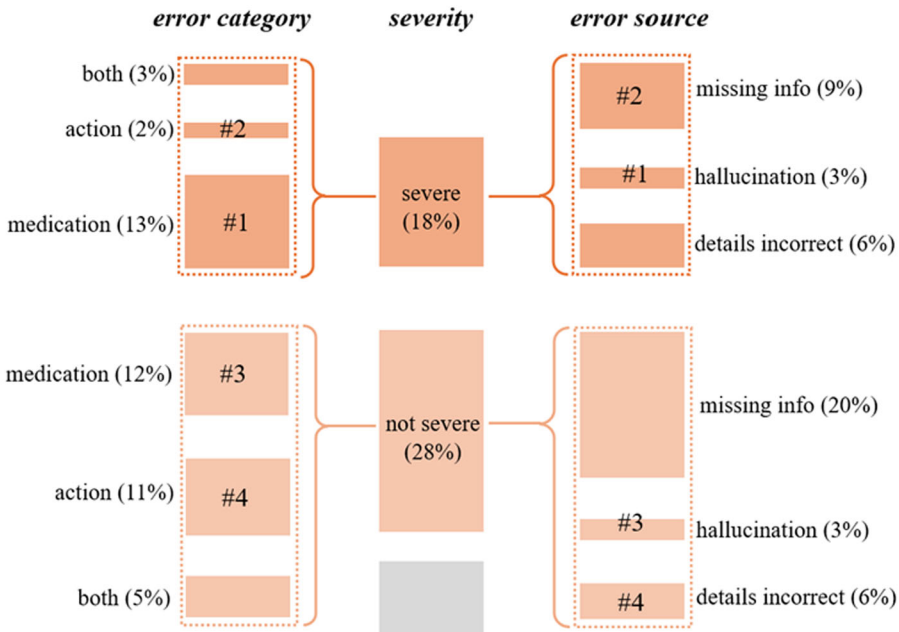
To our knowledge, this study represents the first investigation of the safety, accuracy and language of patient-centred discharge instructions generated from discharge summaries using an AI tool. The results showed that nearly all medications from the original discharge summaries were correctly reflected in the AI-generated responses, though only around half of the follow-up actions were included and new actions were often added. The AI-generated responses were better aligned with health literacy principles than the discharge summaries but only some of the medication instructions could be simplified into a form that is known to be easier for patients to follow. Importantly, potential safety issues were introduced into the instructions.

In related work, the use of generative AI to support the production of discharge summaries has been proposed[22], and early tests for producing discharge summaries have had some positive results[23]. Other research has examined the use of ChatGPT to simplify surgical consent forms[16], and radiology reports[21], as well as general health information available online[17–20]. Others propose generative AI tools as possible solutions to healthcare communication issues but—consistent with our findings—suggest caution in relation to their safety[24–27].

Recent advances in generative AI have enabled the use of general-purpose generative AI tools in clinical workflows, but our findings suggest that more work needs to be done to ensure that the tools are safely adopted in practice and avoid unintended consequences. For producing patient-centred instructions, generative AI could be used to produce a 'first draft' but the need to review the instructions may add to clinical workload. Future research to improve the safety of generative AI used with discharge may benefit from the development and use of tools that help identify the source of safety issues in summarisation tasks, including hallucination, the balance between information extraction and generating new text, or changed meaning through summarisation. Future practice and implementation directions may consider broader goals in transitions of care, including generating discharge summaries from medical records[22,23] and generating multiple discharge documents intended for primary care physicians, patients with different levels of health literacy and culturally and linguistically diverse backgrounds, and other care providers.

This study had several limitations. The set of discharge summaries were from one dataset (MIMIC-IV), which represents one location and results may not generalise to other healthcare systems where discharge summaries differ. Data from MIMIC-IV are de-identified, which means that some details were missing and introduced some ambiguity in the original discharge summaries that occasionally hindered the evaluation of the responses. While we used a robust approach for prompt engineering, evaluations of other prompts may have yielded different results. It may be useful to explicitly separate the information extraction from the summarisation and language simplification as two separate tasks, but this approach may also need to consider how best to incorporate contextual information from other sections of the discharge summary. We used GPT-3.5 as the basis for generating responses, and other language models may also have yielded different results. Future studies in the area could replicate the methods we use here and compare different combinations of language models, prompts and discharge summaries from other locations.

Generative AI tools may be used to support discharge planning by generating new patient-centred discharge instructions, filling an important current gap in communication with patients as they leave hospital. While there is a clear need to improve communication with patients on discharge, AI-generated patient discharge instructions can introduce incorrect information, which in some cases could lead to harm. New language models and advances in prompt engineering may help to balance constraints related to health literacy, accuracy and safety. Before considering the use of AI-generated patient-centred discharge instructions with patients, processes for ensuring safety are needed.



Example 1 (severe, medication error, hallucination)
Original: 17. Fluticasone-Salmeterol Diskus (250/50) 1 INH IH BID; Revised: 17. Fluticasone-Salmeterol Diskus (250/50): Inhale the contents of the diskus once daily for better breathing. Safety issue: Revised texts instructs patient to use the entire diskus rather than a single puff.

Example 2 (severe, action error, missing info)
Original: Pt was instructed to follow up with PCP in next few days re: HTN management. Safety issue: No advice about following up with primary care physician about hypertension management.

Example 3 (not severe, medication error, hallucination)
Original: 6. acetaminophen 325 mg Tablet Sig: ___ Tablets PO Q6H (every 6 hours) as needed for pain: Over the counter. Revised: 5. Acetaminophen 325 mg Tablet: - Take as needed for pain, one tablet every 6 hours. Safety issue: Gives number of tablets when not in the original text.

Example 4 (not severe, action error, details incorrect)
Original: You have dissolvable sutures and may wash your hair and get your incision wet day 3 after surgery. Revised: Wash your hair and get your incision wet on Day 3 after surgery. Safety issue: None. There is a difference in directiveness ('may' vs an instruction)

Fig. 1 | A review of safety issues identified in the AI-generated patient-centred discharge instructions. The safety issues were identified in 100 patient-centred discharge instructions, and descriptions include the severity and provenance of the safety issues. Four examples are identified on the figure based on their error category and source.

Methods

The study design was a comparison between patient-centred discharge instructions generated by prompting a GPT-based model and the doctor-written discharge summaries on which they were based. Evaluations included manual review by experts, and all evaluations of accuracy and safety were undertaken by investigators with qualifications in medicine or pharmacy.

The University of Sydney Research Integrity and Ethics Administration confirmed that the methodology of the study meets ethical review exception guidelines, as per the National Health and Medical Research Council National Statement on Ethical Conduct in Human Research. The study involved the use of existing collections of data or records that contain only non-identifiable data and was deemed to be of negligible risk.

Data sources

Discharge summaries were sourced from the Medical Information Mart for Intensive Care IV (MIMIC-IV) version 2.2 database[28–30]. The database includes deidentified electronic medical records from over 40,000 patients admitted to the Beth Israel Deaconess Medical Centre in Boston, Massachusetts, between 2008 and 2019. All investigators interacting with data from MIMIC-IV were credentialed users of the PhysioNet database. Discharge summaries were randomly sampled from the MIMIC-IV database and used in the development and analysis if they were written in English and if patients were discharged from hospital alive (Supplementary Table 5). Ten discharge summaries were used to help develop prompts and train investigators on the evaluations, and 100 discharge summaries were used in the main evaluation (Supplementary Table 3). Other information from MIMIC-IV related to the patients from the discharge summaries were not accessed or used.

Prompt development and selection

The GPT-3.5 model was accessed via the Microsoft Azure OpenAI service and met the requirements for safe use of MIMIC-IV data. A ChatGPT-like interface was developed to allow the safe access of GPT3.5 to test prompts on examples of discharge summaries from MIMIC-IV (Supplementary Figs. 1–3, Supplementary Boxes 1–3).

The language model takes a prompt and an entire discharge summary as inputs and generates a response. The response is not an extraction of the text in the discharge summary but newly generated text in response to the instructions provided in the prompt. Language models are known to be sensitive to small changes in prompts, so the prompt used in the analysis was developed through a process of iterative refinement and testing.

First, expert-derived examples of patient instructions were created. Five discharge summaries from the MIMIC-IV database were used to derive patient discharge instructions (including medication and action lists) by two investigators. Disagreements were resolved by discussion with the broader group of investigators. Following this step, prompts were iteratively refined and tested to produce responses that most closely

matched five of the expert-derived examples using three prompt design approaches, including 'direct', 'multi-stage' and 'worked example' approaches (Supplementary Figs. 1–3, Supplementary Boxes 1–3). Investigators with clinical expertise scored each of the three prompts across each of five additional examples.

The prompt with the best balance between language complexity and accuracy was selected for the main analysis. The selected prompt was the 'direct' approach, which more often correctly represented medications and included more of the follow-up actions than the other prompts, while still reducing grade reading score and language complexity. Note that a two-step process where information is first extracted from the original discharge summary and then simplified to match the needs of patients may seem like a useful approach. However, the challenge with splitting the approach into two stages starting with information extraction (rather than retrieval augmented generation) is that the whole discharge summary provides contextual information that may be important to the details of the medications and follow-up instructions and direct information extraction would not be able to capture that context in the same way.

Analysis and outcome measures

Each response was independently scored by two investigators with expertise in medicine or pharmacy, comparing each response against the information available in the original discharge summary. Inter-rater reliability scores were calculated using Cohen's Kappa for dichotomous variables and intra-class coefficient for proportional variables. Disagreements were resolved by discussion among the group, producing a final set of scores for each of the 100 discharge summaries. Descriptive statistics were also recorded, including the number of words, medications, and actions in the original discharge summaries and the responses.

Agreement between experts was higher for evaluating whether all discharge medications from the original discharge summary were included in the response (Cohen's kappa 0.889), that no new medications were added (Cohen's kappa 0.852) and the percentage of medications that were presented in UMS format (intra-class correlation coefficient 0.738). Agreement was lower for whether all actions from the original discharge summary were included in the response (Cohen's kappa 0.521), that no new actions were added (Cohen's kappa 0.569), the percentage of medications that were correct (intraclass correlation coefficient 0.438) and the percentage of actions that were correct (intraclass correlation coefficient 0.512).

Clinicians made note of any potential safety issues while evaluating the completeness and accuracy of the medications and follow-up actions, and these notes were discussed as a group to determine severity and provenance. Errors were categorised as errors of omission such as missing instructions, or errors of commission or translation such as a changed dose or route of a medication, inclusion of medications used during a hospital stay and not intended for use after discharge, where a new medication or follow-up action was introduced as a form of hallucination from the AI model.

The accuracy of the AI-generated responses was evaluated using three measures (Table 2). This included whether all medications and actions in the original discharge summary had been included in the patient instructions, whether responses included additional medications or actions that were not present in the post-discharge instructions within the original discharge summary, and the percentage of medications and actions from the original discharge summary that were included and correctly included in terms of dose, route, frequency and duration.

Health literacy was evaluated using three outcome measures (Table 2). Grade reading level and language complexity was measured using the Sydney Health Literacy Lab Health Literacy Editor[24,31]. Grade reading score estimates the level of education that most people would need to correctly understand a given text. The Editor calculates grade reading score using the Simple Measure of Gobbledygook, which is widely used in health literacy research[32]. Language complexity is the percentage of words in the text that are considered medical jargon, acronyms, or uncommon English words.

Table 2 | Study outcome measures and assessment method

Outcome measures	Assessment
Descriptive	
Original discharge summary	
number of words in the original discharge summary	Software
number of medications in the original discharge summary	Expert
number of actions in the original discharge summary, including appointments, tests, behaviours and management plans	Expert
Health literacy	
Language simplification	
grade reading score of the original discharge summary and generated patient discharge instructions	SHeLL Editor
language complexity score of the original discharge summary and generated patient discharge instructions	SHeLL Editor
Universal Medication Schedule (UMS)	
what fraction of medications with a frequency of four or fewer per day were correctly presented in UMS format?	Expert
Accuracy	
Medications	
did the response include all the medications? Y/N	Expert
did the response include any additional medications? Y/N	Expert
what fraction of the medications were correctly specified by type, dose, route, frequency and duration?	Expert
Actions	
did the response include all the action points? Y/N	Expert
did the response include any additional action points? Y/N	Expert
what fraction of the action points were correct?	Expert
Safety	
Potential safety issues	
did the response pass on medication or action information from the original discharge summary in a way that could cause harm? Y/N	Expert
did the response add new information that could cause harm? Y/N	Expert

This calculation was based on existing medical and public health thesauri and an English-language word frequency list. For both measures, lower values correspond to simpler text that should be easier to understand. Paired sample t-tests were used to compare grade reading level and language complexity scores between the original discharge summary and the AI-generated patient-centred discharge instructions. For medications that were prescribed up to four times a day, we manually determined the percentage of medications that were presented in the patient-centred discharge instructions in UMS format.

Data availability

Discharge summaries in MIMIC-IV are available only with permission. We have included the identifiers for the discharge summaries we used in the Supplementary Material (Supplementary Table 5) so researchers with access to MIMIC-IV can replicate the methods or evaluate the impact of using new prompts or language models.

Code availability

The GPT-3.5-turbo-16k 2023-07-01-preview model was freely available online and has been updated (https://platform.openai.com/docs/models/gpt-3-5-turbo). Versions of the lightweight GPT-3.5-turbo can be used in a local implementation of an interface to test new approaches to prompts for safety and language evaluations within the terms and conditions set out for the MIMIC-IV dataset.

Received: 15 January 2024; Accepted: 11 November 2024;
Published online: 20 November 2024

References

1. Wimsett, J., Harper, A. & Jones, P. Review article: components of a good quality discharge summary: a systematic review. *Emerg. Med. Australas.* **26**, 430–438 (2014).
2. Scarfo, N. L. et al. General practitioners' perspectives on discharge summaries from a health network of three hospitals in South Australia. *Aust. Health Rev.* **47**, 433–440 (2023).
3. 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.* **21**, 349 (2021).
4. Mixon, A. S. et al. Characteristics associated with postdischarge medication errors. *Mayo Clin. Proc.* **89**, 1042–1051 (2014).
5. Hoek, A. E. et al. Patient discharge instructions in the emergency department and their effects on comprehension and recall of discharge instructions: a systematic review and meta-analysis. *Ann. Emerg. Med.* **75**, 435–444 (2020).
6. Lindquist, L. A. et al. Relationship of health literacy to intentional and unintentional non-adherence of hospital discharge medications. *J. Gen. Intern. Med.* **27**, 173–178 (2012).
7. Choudhry, A. J. et al. Enhanced readability of discharge summaries decreases provider telephone calls and patient readmissions in the posthospital setting. *Surgery* **165**, 789–794 (2019).
8. Berkman, N. D. et al. Health literacy interventions and outcomes: an updated systematic review. *Evid. Rep. Technol. Assess*. **199**, 1–941 (2011).
9. Sheridan, S. L. et al. Interventions for individuals with low health literacy: a systematic review. *J. Health Commun.* **16**, 30–54 (2011).
10. Brega, A. G. et al. Using the health literacy universal precautions toolkit to improve the quality of patient materials. *J. Health Commun.* **20**, 69–76 (2015).
11. Wolf, M. S. et al. A patient-centered prescription drug label to promote appropriate medication use and adherence. *J. Gen. Intern. Med.* **31**, 1482–1489 (2016).
12. Wolf, M. S. et al. Prevalence of Universal Medication Schedule prescribing and links to adherence. *Am. J. Health Syst. Pharm.* **77**, 196–205 (2020).
13. Wolf, M. S. et al. Effect of standardized, patient-centered label instructions to improve comprehension of prescription drug use. *Med. Care* **49**, 96–100 (2011).
14. Davis, T. C. et al. Improving patient understanding of prescription drug label instructions. *J. Gen. Intern. Med.* **24**, 57–62 (2009).
15. Bailey, S. C., Sarkar, U., Chen, A. H., Schillinger, D. & Wolf, M. S. Evaluation of language concordant, patient-centered drug label instructions. *J. Gen. Intern. Med.* **27**, 1707–1713 (2012).
16. Ali, R. et al. *Bridging the Literacy Gap for Surgical Consents: An AI-Human Expert Collaborative Approach*. http://medrxiv.org/lookup/doi/10.1101/2023.05.06.23289615; https://doi.org/10.1101/2023.05.06.23289615 (2023).
17. Ayre, J. et al. New frontiers in health literacy: using ChatGPT to simplify health information for people in the community. *J. Gen. Intern. Med*. https://doi.org/10.1007/s11606-023-08469-w (2023).
18. Ali, S. R., Dobbs, T. D., Hutchings, H. A. & Whitaker, I. S. Using ChatGPT to write patient clinic letters. *Lancet Digit. Health* **5**, e179–e181 (2023).
19. Ayoub, N. F., Lee, Y.-J., Grimm, D. & Balakrishnan, K. Comparison between ChatGPT and google search as sources of postoperative patient instructions. *JAMA Otolaryngol. Neck Surg.* **149**, 556 (2023).
20. Spallek, S., Birrell, L., Kershaw, S., Devine, E. K. & Thornton, L. Can we use ChatGPT for mental health and substance use education? Examining its quality and potential harms. *JMIR Med. Educ.* **9**, e51243 (2023).
21. Lyu, Q. et al. Translating radiology reports into plain language using ChatGPT and GPT-4 with prompt learning: results, limitations, and potential. *Vis. Comput. Ind. Biomed. Art.* **6**, 9 (2023).
22. Patel, S. B. & Lam, K. ChatGPT: the future of discharge summaries? *Lancet Digit. Health* **5**, e107–e108 (2023).
23. Clough, R. A. et al. Transforming healthcare documentation: harnessing the potential of AI to generate discharge summaries. *BJGP Open* BJGPO.2023.0116 https://doi.org/10.3399/BJGPO.2023.0116 (2023).
24. Ayre, J. et al. Multiple automated health literacy assessments of written health information: development of the SHeLL (Sydney Health Literacy Lab) Health Literacy Editor v1. *JMIR Form. Res.* **7**, e40645 (2023).
25. Dunn, A. G., Shih, I., Ayre, J. & Spallek, H. What generative AI means for trust in health communications. *J. Commun. Healthc.* **16**, 385–388 (2023).
26. Nutbeam, D. Artificial intelligence and health literacy—proceed with caution. *Health Lit. Commun. Open* **1**, 2263355 (2023).
27. Reddy, S., Allan, S., Coghlan, S. & Cooper, P. A governance model for the application of AI in health care. *J. Am. Med. Inform. Assoc.* **27**, 491–497 (2020).
28. Johnson, A. et al. MIMIC-IV. PhysioNet https://doi.org/10.13026/RRGF-XW32.
29. Goldberger, A. L. et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation* **101**, e215–e220 (2000).
30. Johnson, A. E. W. et al. MIMIC-IV, a freely accessible electronic health record dataset. *Sci. Data* **10**, 1 (2023).
31. Sydney Health Literacy Lab. The SHeLL Health Literacy Editor [Webpage].
32. Mc Laughlin, G. H. SMOG grading-a new readability formula. *J. Read*. **12**, 639–646 (1969)

Acknowledgements

No external funding was used for this study.

Author contributions

Study design: K.S., S.Z., X.Z., A.D., J.A.; data extraction and management: K.S., X.Z.; analysis and evaluation: K.S., S.Z., X.Z., S.K., M.T., L.K., S.L., V.N., A.D., J.A.; manuscript development: K.S., S.Z., A.D., J.A.; manuscript critical review and editing: K.S., S.Z., X.Z., S.K., M.T., L.K., S.L., V.N., J.K., A.D., J.A.

Competing interests

J.A. is a director of a health literacy consultancy, which provides health literacy advice to health services and organisations, but no personal income is received. A.D. is Deputy Editor for npj Digital Medicine. All other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41746-024-01336-w.

Correspondence and requests for materials should be addressed to Adam G. Dunn.