On Simplification of Discharge Summaries in Serbian: Facing the Challenges

Anđelka Zečević, Milica Ćulafić, Stefan Stojković
Mathematical Institute, Serbian Academy of Sciences and Arts
Faculty of Pharmacy, University of Belgrade
Clinic for Gastroenterology and Hepatology, University Clinical Center of Serbia, Belgrade
andjelkaz@mi.sanu.ac.rs, milica.culafic@pharmacy.bg.ac.rs, stefanstojkovic@gmail.com

Abstract

The simplified information page (SIP) is a simplified discharge summary created to mitigate health risks caused by low medical comprehension. One of the most critical aspects of medical comprehension concerns interpreting medication instructions such as proper dosing, frequency, and duration. In our work, we examine the capacities of mainstream Large Language Models (LLMs) such as ChatGPT and Gemini to generate SIP-like medication-oriented pages based on the provided discharge summaries. We are sharing the initial qualitative assessments of our study based on a small collection of discharge summaries in Serbian, pointing to noticed inaccuracies, unfaithful content, and language quality. Hopefully, these findings might be helpful in addressing the multilingual perspective of patient-oriented language.

Keywords: patient safety, text simplification, discharge summary, LLMs

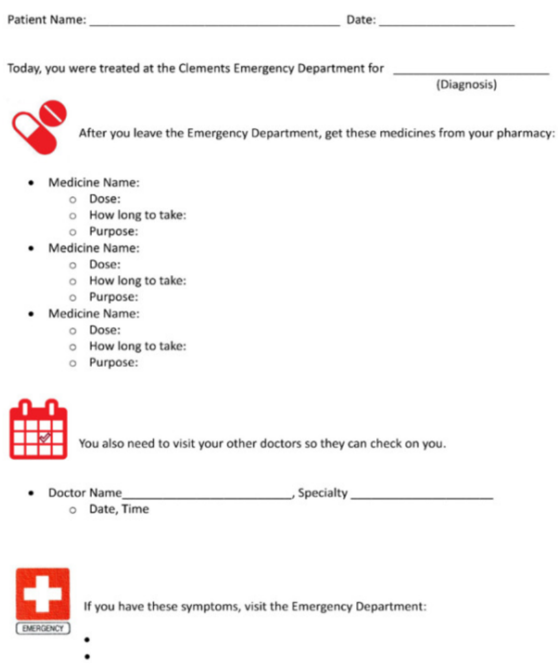
1. Introduction

Understanding discharge summaries is crucial for continuity of care and patient safety. However, low comprehension poses significant challenges in healthcare delivery. Inadequate comprehension of discharge summaries can lead to medication errors, treatment delays, and patient confusion. This is notably well-supported, especially regarding medication errors in post-discharge (Weetman et al., 2021; Alqenae et al., 2020).

Healthcare providers often struggle to communicate complex medical information effectively within discharge summaries, hindering patient understanding. On the other hand, the so-called patient literacy plays a key role in understanding discharge summaries. Regardless of education level, this literacy may be quite limited and patients, thus, face heightened challenges in understanding crucial health-related information. This issue lies at the core of health disparities (Murugesu et al., 2022).

Addressing the challenges associated with a low understanding of discharge summaries requires interdisciplinary efforts involving healthcare providers, policymakers, educators, and technology developers to enhance clarity, accessibility, and patient-centeredness in discharge communication (Geese et al., 2023; Bhati et al., 2023).

One line of research that improves patient understanding is the creation of simplified discharge instructions. The simplified information page (SIP) (DeSai et al., 2021) is a one-page patient discharge summary designed originally for emergency departments in accordance with the Centers for Medicare and Medicaid Services and Joint Commission recommendations. It lists information related to diagnoses, recommended treatments (medications, diet, therapy, wound care, etc.), doctors or clinics needed to follow up, and symptoms or circumstances that should be monitored and urgently addressed (Figure 1). All information are presented in a simplified manner with the Flesch-Kincaid grade level 5. The SIP demonstrates that changing only the information structure and making it more accessible improves patients' comprehension. In the most critical segments that relate to medication dosage and duration, an improvement by over 22% is noticed across all demographics and education levels.


Patient Name: ______________________________ Date: ________________

Today, you were treated at the Clements Emergency Department for ____________________
(Diagnosis)

After you leave the Emergency Department, get these medicines from your pharmacy:

- Medicine Name:
  - Dose:
  - How long to take:
  - Purpose:
- Medicine Name:
  - Dose:
  - How long to take:
  - Purpose:
- Medicine Name:
  - Dose:
  - How long to take:
  - Purpose:

You also need to visit your other doctors so they can check on you.

- Doctor Name____________________, Specialty __________________
  - Date, Time

EMERGENCY

If you have these symptoms, visit the Emergency Department:

-
-

Figure 1: Simplified information page.

Proceedings of CL4Health 2024 @LREC-COLING-2024, pages 104–108
20 May, 2024.

As the manual creation of discharge summaries in the SIP-like form requires additional personnel and time resources, we examine the potential of Large Language Models (LLMs) to recreate them. LLMs have already entered the world of biomedicine with models trained on medical publications (PubMedBERT, Gu et al., 2021), medical records (ClinicalBERT, Huang et al., 2019), or medical knowledge bases (UmlsBert, Kang et al. 2020). The performances on the relevant benchmarks, such as the BLURB - Biomedical Language Understanding and Reasoning Benchmark (Gu et al., 2021), spark the various capabilities of biomedical models. However, the complexity of medicine, the ever-growing medical knowledge, constant technology enhancements, and its safety-critical nature, always reveal the necessity for improvement.

Due to disparities in healthcare digitalization and regulatory policies across different regions, biomedical datasets in non-English languages are often scarce. The same holds for language models and appropriate language tools. Therefore, in our approach, we leverage publicly accessible chatbots such as ChatGPT[1] and Gemini[2], to generate SIPs for the provided expert-written discharge summaries in Serbian. We mainly focus on the medication instruction part, including medication names, dosages, durations, frequencies, ways of administration, and their purpose. Although working with a small collection of discharge summaries, employing a qualitative approach, we were able to identify several pain points of language models that require additional attention and enhancement.

2. Related Work

Soon after progress had been made in language modeling (Vaswani et al., 2017), document summarization began to reflect notable improvements in its ability to distill key information from large volumes of text (Liu and Lapata, 2019; Lewis et al., 2019; Raffel et al., 2020), mostly in the general domain. Medical document summarization is, however, somewhat different as it poses several challenges, including handling complex medical terminology, high accuracy expectations, and preserving patient privacy and confidentiality. What come as natural tasks are the summarization of medical notes (Landes et al., 2023), medical research (Devaraj et al., 2021; Singhal et al., 2023) as well doctor-patient conversations (Abacha et al., 2023).

From the perspective of a clinician, a discharge summary represents a concise overview of the patient's course of hospitalization, treatment, and follow-up care plan that can serve as a communication tool that facilitates continuity of care between the hospital and outpatient settings. It also represents a demanding, time-consuming administrative activity based on abundant medical documentation that is oftentimes challenging to digest. Therefore, the existing work mostly tries to alleviate this setup (Shing et al., 2021; Searle et al., 2023), the latest one being the *Discharge Me!*[3], a BioNLP ACL'24 Shared Task on Streamlining Discharge Documentation.

Although there are publications addressing the patient aspect of medical summarization (Zaretsky et al., 2024), they are less present. We hope that our work can help fill in the gap by combining the imperatives of both sides into a unified goal.

3. Experiment

As stated, our goal was to examine the capacity of publicly available mainstream LLMs to generate SIP-like medication-oriented lists easily accessible by patients. For that purpose, we collected a number of discharge summaries in Serbian, prompted ChatGPT and Gemini to generate SIPs, and manually evaluated the results we obtained.

3.1 Dataset

We started our work with a small collection of discharge summaries in Serbian, in total 13, provided by the Liver Transplant Unit of the Clinic for Gastroenterology and Hepatology at the University Clinical Centre of Serbia. All discharge summaries are read by one medical professional and anonymized according to the local privacy regulations by masking patient-related information, dates, ambulance names, names of practicing doctors, names of doctors who are meant to perform additional examinations, and phone numbers for scheduling examinations and obtaining information.

Due to the complexity of cases, discharge summaries were very diverse in terms of medication instructions. In total, 65 medications are covered, of which 38 are unique, including different dosage forms (tablets, capsules, droplets, sprays) and routes of administration. The average number of medications per discharge summary was 5.

The average length of the section of discharge summaries comprising prescribed medications and follow-up care instructions was 97 tokens, indicating short and condensed directions to the patients. We used Latin script as it was used in the original discharge summaries. We did not use any preprocessing step prior to utilizing language models.

[1] https://chat.openai.com/

[2] https://gemini.google.com/

[3] https://stanford-aimi.github.io/discharge-me/

3.2 Prompts

To generate SIP-like lists, we prompted the models with the template written in Serbian in a zero-shot manner. The appropriate prompt translation in English is given below.

I will forward the patient's discharge summary from the Clinic for Gastroenterology and Hepatology. You should single out each medication, its dosage, its method of administration, its frequency, and a short explanation of what the drug is used for. In addition, single out notes related to further examinations or controls. If abbreviations appear in the result, please provide the corresponding meanings.

Figure 2: Initial prompt translated into English.

The medication-related information such as medication name, dosage, and frequency was part of the discharge summary and easily accessible to LLMs. The method of administration depended on the medication form and was supposed to be concluded by LLMs. The same held for medication purposes and short descriptions that were to be generated based on LLMs' medical knowledge.

In cases where medication instructions vary for different days of the week, we prompted LLMs additionally for day-dependant SIP-like lists by utilizing the template below.

Can you now create a list with appropriate medications for each day of the week?

Figure 3: Day-dependant prompt translated into English.

4. Results

For each discharge summary, we prompted ChatGPT and Gemini using prepared templates with the goal of generating a medication-oriented SIP-like list. All results were manually evaluated by one medical expert by carefully comparing the original discharge summary and generated SIP lists. For each medication, the evaluator scored if the medication was present on the SIP list, if dosage, dosage form, route of administration, its frequency, and duration (when stated) were appropriate, and if a short description of the medication's purpose was correct. The total number of evaluated medications in the dataset was 65. Table 1. summarizes our main quantitative findings related to medication inconsistencies.

Both models correctly extracted medication names from discharge summaries. The exceptions were medications *Entyvio* and *Zometa*, not directly prescribed by the doctor but mentioned as a part of the patient's existing medication protocol. However, not all medication descriptions and purposes were appropriate. For example, ChatGPT explained that *Oglition* is a cholesterol-lowering medication, while Gemini explained it is an immunosuppressor. None of these is correct, as *Oglition* is primarily used as an antidiabetic. In order to validate the claims, we relied on the expert opinion and package leaflets available on the official website of the Agency for Medicines and Medical Devices of Serbia.

	ChatGPT	Gemini
Medication omission	1	2
Inappropriate description	2	7
Inappropriate frequency	9	11

Table 1: The type and frequency of noticed inconsistencies between original discharge summaries and SIPs. The total number of revised medications is 65.

As precise dosages and dosage forms were present for each medication in discharge summaries (for example, *Advagraf caps. a 1mg 1x2*), we did not record any inconsistencies related to these parts. However, instructions related to medication frequency were the most challenging for the models to interpret and verbalize. Within discharge summaries, there were two different ways of specifying frequencies: *frequency x dosage* and *morning dosage + noon dosage + evening dosage*. Simple instructions, such as *Pravacor tabl. a 20mg 1x1* containing *1x1* form, were successfully interpreted in all cases. Instructions containing forms such as *1x2* or *2x1* often led to swapping the frequency and dosage in the generated narratives. For example, instruction *Imuran tabl. a 50mg 1x3* was interpreted as taking one tablet three times a day, every 8 hours, instead of taking one dose (consisting) of three tablets. The interpretation was even less successful in the case of fractions, for example, with frequency forms such as *1x1/2* or *2x1/4*. Gemini could not interpret these instructions at all, as denominators were excluded from the generated descriptions. Therefore, the instruction such as *Propranolol tabl. a 40mg 2x1/4* was interpreted as *Propranolol tabl. a 40mg 2x1,* leading to a much higher dosage. ChatGPT was partially successful but inconsistent within sessions. Instructions in the form *morning dosage + noon dosage + evening dosage* were correctly interpreted by both models.

Frequency instructions, such as *three times a week* or with an explicit list of the days (Monday/Wednesday/Friday) were correctly extracted by the models but partially utilized. For example, when prompted to generate SIP lists for each day of the week, *Vigantol* droplets originally prescribed with the instruction *10 drops three times a week*, were repeated for all days by ChatGPT. Gemini could interpret this instruction correctly and even visualize the table with weekday names as headers.

Through the examples, we noticed that models can point to unspecified medical instructions. For example, the instruction to take medicine at 8h was ambiguous, as it was unclear if 8h relates to the morning or evening hours. We found these rare cases of ambiguity particularly important as they can also cause patients to feel unsure and hesitant to act.

Both models demonstrated accuracy in extracting information concerning future appointments, primarily pertaining to additional analyses, biopsies, and scans. They proved particularly useful in clarifying abbreviations associated with medical procedures and dietary regimens.

The content generated by ChatGPT was grammatical and of satisfactory quality. On the other hand, Gemini often code-switched between Serbian and English and even mixed Latin and Cyrillic script. This required additional post-processing and content validation.

When prompted several times within one session, ChatGPT started combining discharge summaries and generating improper, hallucinated, medications. Therefore, each experiment was performed within a unique session. We did not notice similar behavior while using Gemini.

5. Conclusion

Our study aims to improve communication between healthcare professionals and their patients. Simplified discharge summaries should translate complex medical information into a more comprehensive language, which should positively impact patient literacy in general. Patient literacy is an ever-growing concept medical experts use to denote levels by which individuals perceive or learn to comprehend health information within the decision-making processes. In confronting the existing challenges concerning the low understanding of discharge summaries, we qualitatively analyze a small dataset in the Serbian language - a language not equally covered in international protocols. Therefore, disseminating these summaries, knowledge gaps, and improvements will lead us to relevant statistics and possible solutions. Hopefully, these results will encourage scholars, stakeholders, and members of the healthcare system to strive to find more accessible paths for delivering better quality care.

Presently accessible general-purpose LLMs exhibit promise in producing simplified discharge summaries, even for languages with limited resources like Serbian. Nonetheless, these summaries do not consistently align with the original summaries concerning critical medical elements such as medication frequency or purpose, thereby compromising their reliability.

As shared results represent only a fraction of our ongoing research, the list of forthcoming activities needed for deeper validation and improvements is extensive. We plan to experiment with a larger dataset that includes other medical subfields and discharge summary writing styles. In order to alleviate the influence of prompts, we plan to design and perform additional behavior consistency experiments. Further, we plan to prepare a supporting dataset for the training of RAG architecture (Lewis et al., 2020) with the aspiration of addressing the previously highlighted accuracy and trustworthy-related observations. Finally, we plan to conduct an on-site evaluation of SIP lists with patients and medical personnel to obtain a qualitative assessment of the proposed methodology in terms of improved medical literacy and instruction comprehension.

6. Bibliographical References

Abacha, Asma Ben, Wen-wai Yim, Griffin Adams, Neal Snider, and Meliha Yetisgen. (2023). *Overview of the MEDIQA-Chat 2023 Shared Tasks on the Summarization & Generation of Doctor-Patient Conversations*. In *Proceedings of the 5th Clinical Natural Language Processing Workshop*, pages 503–513, Toronto, Canada. Association for Computational Linguistics.

Alqenae F. A., Steinke, D., Keers, R. N. (2020). *Prevalence and nature of medication errors and medication-related harm following discharge from hospital to community settings: a systematic review.* Drug Saf, 43(6):517–537.

Bhati D., Deogade M. S., Kanyal D. (2023). *Improving Patient Outcomes Through Effective Hospital Administration: A Comprehensive Review.* Cureus, 15(10):e47731.

DeSai, C., Janowiak, K., Secheli, B., Phelps, E., McDonald, S., Reed, G., & Blomkalns, A. (2021). Empowering patients: Simplifying discharge instructions. *BMJ Open Quality*, *10*(3).

Devaraj, Ashwin, Iain Marshall, Byron Wallace, and Junyi Jessy Li. (2021). *Paragraph-level Simplification of Medical Texts*. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 4972–4984.

Geese, F., Schmitt, K. U. (2023). *Interprofessional Collaboration in Complex Patient Care Transition: A Qualitative Multi-Perspective Analysis.* Healthcare, 11, 359.

Gu, Yu, Robert Tinn, Hao Cheng, Michael Lucas, Naoto Usuyama, Xiaodong Liu, Tristan Naumann, Jianfeng Gao, and Hoifung Poon. (2021). *Domain-specific language model pretraining for biomedical natural language processing*. *ACM Transactions on Computing for Healthcare (HEALTH)* 3, no. 1: 1-23.

Huang, Kexin, Jaan Altosaar, and Rajesh Ranganath. (2019). *Clinicalbert: Modeling clinical notes and predicting hospital readmission. arXiv preprint arXiv:1904.05342*.

Kang, Tian, Adler Perotte, Youlan Tang, Casey Ta, and Chunhua Weng. (2021). *UMLS-based data augmentation for natural language processing of clinical research literature. Journal of the American Medical Informatics Association* 28, no. 4: 812-823.

Lewis, Mike, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. (2019). *Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension*. *arXiv preprint arXiv:1910.13461*.

Lewis, Patrick, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler et al. (2020). *Retrieval-augmented generation for knowledge-intensive NLP tasks*. *Advances in Neural Information Processing Systems* 33: 9459-9474.

Murugesu L, Heijmans M., Rademakers J., Fransen M. P. (2022). *Challenges and solutions in communication with patients with low health literacy: Perspectives of healthcare providers.* PLoS One, 17(5):e0267782.

Paul Landes, Aaron Chaise, Kunal Patel, Sean Huang, and Barbara Di Eugenio. (2023). *Hospital Discharge Summarization Data Provenance*. In *The 22nd Workshop on Biomedical Natural Language Processing and BioNLP Shared Tasks*, pages 439–448, Toronto, Canada. Association for Computational Linguistics.

Raffel, Colin, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. (2020). *Exploring the limits of transfer learning with a unified text-to-text transformer*. Journal of Machine Learning Research 21, no. 140: 1-67.

Searle, T., Ibrahim, Z., Teo, J. and Dobson, R.J., (2023). *Discharge summary hospital course summarisation of in patient Electronic Health Record text with clinical concept guided deep pre-trained Transformer models*. *Journal of Biomedical Informatics*, *141*, p.104358.

Shaib, C., Li, M.L., Joseph, S., Marshall, I.J., Li, J.J. and Wallace, B.C., (2023). Summarizing, simplifying, and synthesizing medical evidence using GPT-3 (with varying success). *arXiv preprint arXiv:2305.06299*.

Shing, H.C., Shivade, C., Pourdamghani, N., Nan, F., Resnik, P., Oard, D. and Bhatia, P., (2021). *Towards clinical encounter summarization: Learning to compose discharge summaries from prior notes*. *arXiv preprint arXiv:2104.13498*.

Singhal, Karan, Shekoofeh Azizi, Tao Tu, S. S. Mahdavi, Jason Wei, Hyung W. Chung, Nathan Scales et al. (2023). *Large Language Models Encode Clinical Knowledge*. *Nature620*, no. 7972: 172-180.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I. (2017). *Attention is all you need*. *Advances in neural information processing systems*, *30*.

Weetman, K., Spencer, R., Dale, J. et al. (2021). *What makes a "successful" or "unsuccessful" discharge letter? Hospital clinician and General Practitioner assessments of the quality of discharge letters.* BMC Health Serv Res. 2021 Apr 15;21(1):349.

Zaretsky J, Kim JM, Baskharoun S, et al. (2024). *Generative Artificial Intelligence to Transform Inpatient Discharge Summaries to Patient-Friendly Language and Format*. *JAMA Netw Open.* 7(3):e240357.