LREC-COLING 2024

First Workshop on Patient-Oriented Language Processing @LREC-COLING-2024 (CL4Health)

Workshop Proceedings

Editors
Dina Demner-Fushman, Sophia Ananiadou,
Paul Thompson and Brian Ondov

20 May, 2024
Torino, Italia

Proceedings of the First Workshop on Patient-Oriented Language Processing @LREC-COLING-2024 (CL4Health)

ISBN 978-2-493814-25-8
ISSN 2951-2093 (COLING); 2522-2686 (LREC)

Jointly organized by the ELRA Language Resources Association
and the International Committee on Computational Linguistics

Preface

The First Workshop on Patient-Oriented Language Processing (CL4Health) aims to establish a general venue for presenting research and applications focused on patients' needs. These include summarizing health records for patients, answering consumer-health questions using reliable resources, detecting misinformation or potentially harmful information, and providing multi-modal information, such as video, if it better satisfies patients' needs. Such a venue is needed both to invigorate patient-oriented language processing research and to build a community of researchers interested in this area. The growing interest in this topic is fueled by several current trends, which include a proliferation of online services that target patients but do not always act in their best interests; policy changes that allow patients to access their health records written in the professional vernacular, which may confuse the patients or lead to misinterpretation; replacement of customer services with chat bots; and the increasing tendency of patients to consult online resources as a second or even first opinion on their health problems.

Broadly, CL4Health is concerned with the resources, computational approaches, and behavioral and socio-economic aspects of the public interactions with digital resources in search of health-related information that satisfies their information needs and guides their actions.

Invited Speakers

The invited speakers have devoted significant parts of their research to patient-centered language processing. We are grateful and excited to present the following talks:

Barbara Di Eugenio, University of Illinois Chicago, USA

Engaging the Patient in Healthcare: Summarization and Interaction

Effective and compassionate communication with patients is becoming central to healthcare. The talk discusses the results of and lessons learned from three ongoing projects in this space. The first, MyPHA, aims to provide patients with a clear and understandable summary of their hospital stay, which is informed by doctors' and nurses' perspectives, and by the strengths and concerns of the patients themselves. The second, VIRTUAL-COACH, models health coaching interactions via text exchanges that encourage patients to adopt specific and realistic physical activity goals. The third, HFChat, envisions an always-on-call conversational assistant for heart failure patients, that they can ask for information about lifestyle issues such as food and exercise.

Brief Biography

Dr. Di Eugenio's work is characterized by: large interdisciplinary groups of investigators who bring different perspectives to the research; grounding computational models in ecologically valid data, which is small by its own nature; and the need for culturally valid interventions, since the University of Illinois Health system predominantly serves underprivileged, minority populations.

Natalia Grabar, University of Lille, France

Linguistic Foundations of the Simplification and its Current State

The purpose of text simplification is to adapt the content of documents in order to make their reading and understanding easier for a given type of population. If the simplification usually aims specific language levels (lexical, morphological, syntactic, semantic...), the available data cannot always provide precise indications required for this process. The talk discusses some sources of such available data. Dr. Grabar also analyzes the current situation related to the exploitation of linguistic indicators during the definition of language complexity and the simplification.

Brief Biography

Dr. Grabar is a CNRS Researcher at the University of Lille. She studied philology at Lviv University, Ukraine and obtained her PhD in Medical Informatics from the Université Paris 6, France. She develops linguistic and statistical methods to access information and knowledge within scientific and technical texts and terminologies. The results are used in information retrieval, information extraction and text simplification. Dr. Grabar has co-authored over 200 publications.

Graciela Gonzalez-Hernandez, Cedars Sinai Medical Center, USA

Patients are speaking - are we listening? Incorporating patient perspectives posted online into clinical trials

Research that aims to be equitable and effective at treating chronic diseases and improving patient outcomes must incorporate a broad range of patient perspectives (health-related uncertainties, beliefs, and experiences). Setting research priorities and designing trials is complex since clinicians, researchers, and patients differ on what is considered important. Patients often prioritize outcomes that directly impact their quality of life, such as symptom relief, functional status, and treatment side effects, while clinicians prioritize outcomes related to survival, disease progression, and biomarker endpoints. Methods commonly used for gaining patient perspectives are often limited are subject to recall and other biases, are expensive and time-consuming, are limited in recruitment number and diversity, and may not comprehensively capture factors important for research design.

A vast amount of data from the patient's perspective is already publicly available: patients openly share useful perspectives on different social media platforms. Despite its potential, approaches for the systematic integration of such data to inform the prioritization and design of health research are still to be developed and validated.

In this talk, Prof. Gonzalez-Hernandez discusses her ongoing efforts to enable the extraction of relevant patient perspectives posted online using state-of-the-art natural language processing (NLP) methods, and the promise of their integration into clinical trial design.

Brief Biography

Dr. Gonzalez-Hernandez has over 23 years of experience and more than 200 publications in health AI and NLP, funded by multiple NIH grants. She is currently a Professor and Vice Chair for Research and Education in the Cedars-Sinai Department of Computational Biomedicine. She launched the #SMM4H (Social Media Mining for Health) Workshop and Shared Tasks, which has run annually for the last 8 years.

Abeed Sarker, Emory School of Medicine, USA

Learning and Educating via NLP of Social Media: the Use Case for Substance Use and Overdose in the United States

Substance use and overdose is an ongoing crisis in the United States and growing globally. The sphere of substance-related overdose also evolves continuously as novel psychoactive substances enter the supply. Nonmedical substance use surveillance via social media has the potential to provide low-cost and more timely insights than traditional approaches. In our research, we leverage natural language processing (NLP) and machine learning to obtain insights from targeted cohorts of people who use substances about emerging patterns and problems in substance use disorder and treatment. This talk outlines our NLP pipeline for analyzing substance use-related chatter from Twitter (X) and Reddit, and how insights derived from these sources may be used to educate medical practitioners at the forefront of the opioid crisis in the United States, facilitating more patient-centered care.

Brief Biography

Dr. Sarker is an Associate Professor and the Vice Chair for Research at the Department of Biomedical Informatics, School of Medicine, Emory University. He leads several large-scale projects focusing on the application of NLP for health-related tasks, particularly those involving vulnerable populations such as people with substance use disorders, victims of intimate partner violence, and people at risk of self-harm and suicide. His research is primarily funded by the National Institutes of Health (NIH) and Centers for Disease Control and Prevention (CDC). Dr. Sarker's research has been covered by various national and international media outlets such as the Wall Street Journal, Forbes, and Scripps National News.

Submissions

CL4Health received 40 valid submissions, of which 8 were accepted as oral presentations and 25 as posters. The work covers a wide range of topics focusing on patients' well-being and proper care. The topics include retrieval augmented generation, communications (including plain language, sign language, and dialog), mental health issues, and patients' sentiment.

As always, we are deeply grateful to the authors of the submitted papers and to the reviewers (listed elsewhere in this volume) who produced thorough and thoughtful reviews for each paper in a fairly short review period. The Organizers are truly grateful to our amazing Program Committee, whose members helped us determine which studies are ready to be presented

and those which would benefit from additional experiments and analysis, as suggested by the reviewers. We hope that this workshop will inspire new collaborations and research into patient-centered language technologies, in order to continue the valuable contributions made by our community towards public health and well-being.

Dina Demner-Fushman, Sophia Ananiadou, Paul Thompson and Brian Ondov (Organizers)

Organizing Committee

Dina Demner-Fushman, National Library of Medicine, USA
Sophia Ananiadou, National Centre for Text Mining and University of Manchester, UK
Paul Thompson, National Centre for Text Mining and University of Manchester, UK
Brian Ondov, National Library of Medicine, USA

Program Committee

Sophia Ananiadou, National Centre for Text Mining and University of Manchester, UK
Luiz Henrique Bonifacio, University of Waterloo, Canada
Leonardo Campillos-Llanos, Spanish National Research Council, Spain
Dina Demner-Fushman, National Library of Medicine, USA
Manas Gaur, University of Maryland, Baltimore County, USA
Natalia Grabar, Université de Lille, France
Cyril Grouin, Université de Paris-Saclay, CNRS, LISN, Orsay, France
Tudor Groza, Curtin University, Australia
Deepak Gupta, National Library of Medicine, USA
Anna Koroleva, Springbok AI, UK
Alberto Lavelli, Fondazione Bruno Kessler, Italy
Aurélie Névéol, Université de Paris-Saclay, CNRS, LISN, Orsay, France
Brian Ondov, National Library of Medicine, USA
Anthony Rios, University of Texas at San Antonio, USA
Miguel Rocha, University of Minho, Portugal
Roland Roller, German Research Center for Artificial Intelligence, DFKI, Germany
Abeed Sarker, Emory School of Medicine, USA
Sarvesh Soni, National Library of Medicine, USA
Paul Thompson, National Centre for Text Mining and University of Manchester, UK
Grigorios Tsoumakas, Aristotle University of Thessaloniki, Greece
Aswathy Velutharambath, University of Stuttgart, Germany
Amelie Wührl, University of Stuttgart, Germany
Pierre Zweigenbaum, Université de Paris-Saclay, CNRS, LISN, Orsay, France

Invited Speakers

Barbara Di Eugenio, University of Illinois Chicago, USA
Graciela Gonzalez-Hernandez, Cedars Sinai Medical Center, USA
Natalia Grabar, Université de Lille, France
Abeed Sarker, Emory School of Medicine, USA

Table of Contents

Workshop Program

Monday May 20, 2024

09:00–09:05 **Opening remarks**

09:05–10:30 **Session 1: Communicating with patients**

09:05–09:35 ***Invited talk – Barbara Di Eugenio: Engaging the Patient in Healthcare: Summarization and Interaction***

09:35–09:55 *Improving Sign Language Production in the Healthcare Domain Using UMLS and Multi-task Learning*
Jonathan David Mutal, Raphael Rubino, Pierrette Bouillon, Bastien David, Johanna Gerlach and Irene Strasly

09:55–10:15 *It's Difficult to Be Neutral – Human and LLM-based Sentiment Annotation of Patient Comments*
Petter Mæhlum, David Samuel, Rebecka Maria Norman, Elma Jelin, Øyvind Andresen Bjertnæs, Lilja Øvrelid and Erik Velldal

10:15–10:30 ***Poster boasters***

10:30–11:00 ***Coffee break***

11:00–13:00 **Session 2: Patients' language and care**

11:00–11:30 ***Invited talk – Natalia Grabar: Linguistic Foundations of the Simplification and its Current State***

11:30–11:50 *Simulating Diverse Patient Populations Using Patient Vignettes and Large Language Models*
Daniel Reichenpfader and Kerstin Denecke

11:50–12:10 *Annotating Emotions in Acquired Brain Injury Patients' Narratives*
Salomé Klein, Amalia Todirascu, Hélène Vassiliadou, Marie Kuppelin, Joffrey Becart, Thalassio Briand, Clara Coridon, Francine Gerhard-Krait, Joé Laroche, Jean Ulrich and Agata Krasny-Pacini

Monday May 20, 2024 (continued)

12:10–12:30 *Structuring Clinical Notes of Italian ST-elevation Myocardial Infarction Patients*
Vittorio Torri, Sara Mazzucato, Stefano Dalmiani, Umberto Paradossi, Claudio Passino, Sara Moccia, Silvestro Micera and Francesca Ieva

12:30–13:00 ***Poster boasters***

13:00–14:30 ***Lunch***

14:30–16:30 **Poster session (parallel)**

Towards AI-supported Health Communication in Plain Language: Evaluating Intralingual Machine Translation of Medical Texts
Silvana Deilen, Ekaterina Lapshinova-Koltunski, Sergio Hernández Garrido, Christiane Maaß, Julian Hörner, Vanessa Theel and Sophie Ziemer

Large Language Models as Drug Information Providers for Patients
Luca Giordano and Maria Pia di Buono

Towards Generation of Personalised Health Intervention Messages
Clara Wan Ching Ho and Volha Petukhova

Analysing Emotions in Cancer Narratives: A Corpus-Driven Approach
Daisy Monika Lal, Paul Rayson, Sheila A. Payne and Yufeng Liu

Study of Medical Text Reading and Comprehension through Eye-Tracking Fixations
Oksana Ivchenko and Natalia Grabar

A Neuro-Symbolic Approach to Monitoring Salt Content in Food
Anuja Tayal, Barbara Di Eugenio, Devika Salunke, Andrew D. Boyd, Carolyn A. Dickens, Eulalia P. Abril, Olga Garcia-Bedoya and Paula G. Allen-Meares

On Simplification of Discharge Summaries in Serbian: Facing the Challenges
Anđelka Zečević, Milica Ćulafić and Stefan Stojković

Medical-FLAVORS: A Figurative Language and Vocabulary Open Repository for Spanish in the Medical Domain
Lucia Pitarch, Emma Angles-Herrero, Yufeng Liu, Daisy Monika Lal, Jorge Gracia, Paul Rayson and Judith Rietjens

Monday May 20, 2024 (continued)

14:30–16:30 Virtual poster session (parallel)

Revisiting the MIMIC-IV Benchmark: Experiments Using Language Models for Electronic Health Records
Jesus Lovon-Melgarejo, Thouria Ben-Haddi, Jules Di Scala, Jose G. Moreno and Lynda Tamine Lechani

Unraveling Clinical Insights: A Lightweight and Interpretable Approach for Multimodal and Multilingual Knowledge Integration
Kanimozhi Uma and Marie-Francine Moens

Automated Question-Answer Generation for Evaluating RAG-based Chatbots
Juan José González Torres, Mihai Bogdan Bîndilă, Sebastiaan Hofstee, Daniel Szondy, Quang-Hung Nguyen, Shenghui Wang and Gwenn Englebienne

Speech Accommodation in Health-Care Interactions: Evidence Using a Mixed-Reality Platform
Rose Baker, Susan C. Bobb, Dai'Sha Dowson, Elisha Eanes, Makyah McNeill, Hannah Ragsdale, Audrey Eaves, Joseph G. Lee and Kathrin Rothermich

Enhancing Consumer Health Question Reformulation: Chain-of-Thought Prompting Integrating Focus, Type, and User Knowledge Level
Jooyeon Lee, Luan Huy Pham and Özlem Uzuner

Exploring the Challenges of Behaviour Change Language Classification: A Study on Semi-Supervised Learning and the Impact of Pseudo-Labelled Data
Selina Meyer, Marcos Fernandez-Pichel, David Elsweiler and David E. Losada

Using BART to Automatically Generate Discharge Summaries from Swedish Clinical Text
Nils Berg and Hercules Dalianis

Development of a Benchmark Corpus for Medical Device Adverse Event Detection
Susmitha Wunnava, David A. Harris, Florence T. Bourgeois and Timothy A. Miller

16:00–16:30 ***Coffee break***

Monday May 20, 2024 (continued)

16:30–18:00 Session 3: Social media and literature

16:30–17:00 ***Invited talk – Abeed Sarker: Learning and Educating via NLP of Social Media: the Use Case for Substance Use and Overdose in the United States***

17:00–17:20 *Biomedical Entity Linking for Dutch: Fine-tuning a Self-alignment BERT Model on an Automatically Generated Wikipedia Corpus*
Fons Hartendorp, Tom Seinen, Erik van Mulligen and Suzan Verberne

17:20–17:40 *Unveiling Voices: Identification of Concerns in a Social Media Breast Cancer Cohort via Natural Language Processing*
Swati Rajwal, Avinash Kumar Pandey, Zhishuo Han and Abeed Sarker

17:40–18:00 *Intent Detection and Entity Extraction from Biomedical Literature*
Ankan Mullick, Mukur Gupta and Pawan Goyal

18:00–18:05 Closing remarks

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

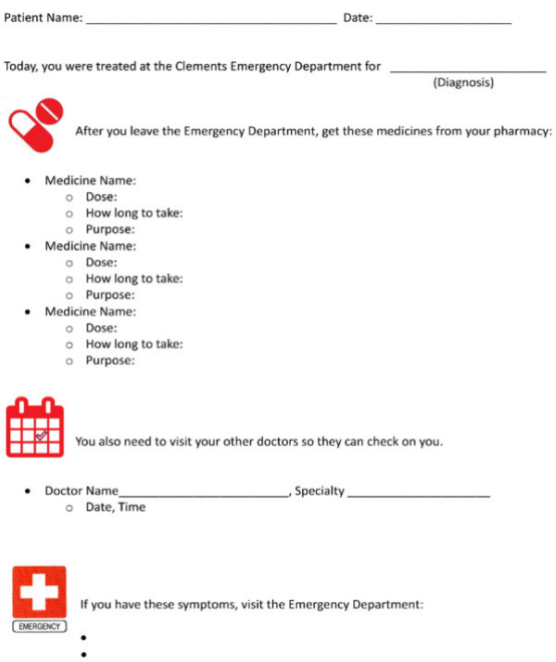
1. Introduction

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

If you have these symptoms, visit the Emergency Department:

EMERGENCY

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.