IgnitionInnovators at "Discharge Me!": Chain-of-Thought Instruction Finetuning Large Language Models for Discharge Summaries

An Quang Tang
RMIT University, Australia
s3695273@rmit.edu.vn

Xiuzhen Zhang
RMIT University, Australia
xiuzhen.zhang@rmit.edu.au

Minh Ngoc Dinh
RMIT University, Australia
minh.dinh4@rmit.edu.vn

Abstract

This paper presents our proposed approach to the Discharge Me! shared task, collocated with the 23th Workshop on Biomedical Natural Language Processing (BioNLP). In this work, we develop an LLM-based framework for solving the Discharge Summary Documentation (DSD) task, i.e., generating the two critical target sections 'Brief Hospital Course' and 'Discharge Instructions' in the discharge summary. By streamlining the recent instruction-finetuning process on LLMs, we explore several prompting strategies for optimally adapting LLMs to specific generation task of DSD. Experimental results show that providing a clear output structure, complimented by a set of comprehensive Chain-of-Thoughts (CoT) questions, effectively improves the model's reasoning capability, and thereby, enhancing the structural correctness and faithfulness of clinical information in the generated text. Source code is available at: `https://anonymous.4open.science/r/Discharge_LLM-A233`

1 Introduction

Discharge summaries encapsulate key details of a patient's hospitalization, from admission to discharge. These documents, however, can contain excessive amount of medical notes, making it difficult for subsequent caregivers or patients to quickly understand essential past medical information. *Brief Hospital Course* and *Discharge Instructions* then become two critical sections in discharge summaries to address this issue. The former outlines critical hospital events for healthcare providers, while the latter offers post-discharge care instructions to patients and their caregivers. The Discharge Me! shared task [1] (Xu et al., 2024) at the BioNLP Workshop, known as Discharge Summary Documentation (DSD), focuses on efficiently generating these two critical sections, a task that is both challenging and time-consuming for clinicians.

In this paper, we introduce a novel LLM-based framework, namely Discharge-LLM, for the DSD task (Xu et al., 2024). Discharge-LLM employs modern prompting strategies (e.g., Chain-of-Thought (CoT)) into instruction-finetuning a `Mistral` Large Language Model (LLM), which enhances structural correctness and faithfulness of clinical information to generate the Brief Hospital Course and Discharge Instructions sections of discharge summaries.

2 Related Work

In recent years, Large Language Models (LLMs) like GPT-2 or GPT-3 have excelled in NLP tasks such as language generation, question answering, due to their vast number of paramters and extensive training on diverse datasets. These models can be adapted to new domains and tasks through methods like prompting, which uses natural language instructions (Liu et al., 2023) or few-shot examples (Lampinen et al., 2022). However, considering DSD problem, the length and excessive information in discharge summaries hinders their use as examples for few-shot prompting. Alternatively, parameter-efficient fine-tuning, which freezes an LLM weights and inserts a small number of tunable parameters (Lin et al., 2020), has proven effective in specialized clinical tasks like radiology report generation (Van Veen et al., 2023).

From the clinical summarization perspective, research towards th DSD task was very limited. But there has been a growing focus on many clinical text generation tasks, encompassing radiology reports (Ben Abacha et al., 2021), clinical notes (Grambow et al., 2022), and summary of diagnoses and patient problems (Gao et al., 2022).

3 Problem description

A discharge summary can contain several free-text sections and medical notes compiled from EHR.

[1] `https://www.codabench.org/competitions/2008/`

Proceedings of the 23rd Workshop on Biomedical Language Processing, pages 731–739
August 16, 2024. ©2024 Association for Computational Linguistics

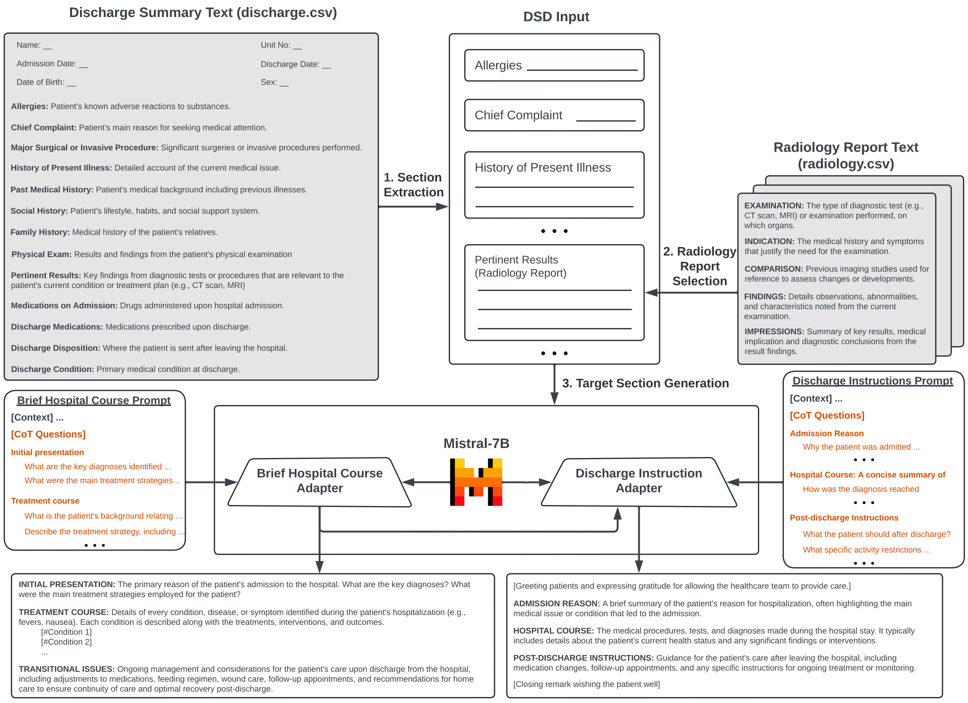


Figure 1: The Discharge-LLM framework

The task is to generate the *Brief Hospital Course* (BHC) and *Discharge Instructions* (DI) sections, leveraging readily available data in other sections of discharge summaries and additional information about a patient's admission (e.g.,radiology, stays) from the dataset. Brief Hospital Course outlines critical hospital events for healthcare providers, while Discharge Instructions offers post-discharge care instructions to patients and their caregivers.

4 Methodology

We propose the Discharge-LLM framework, which adapt LLM to the each generation task of DSD, illustrated in Figure 1. Discharge-LLM applies three steps, namely *Section Extraction*, *Radiology Report Selection* and *Target Section Generation* to generate the two critical target sections given discharge summary and radiology report information of a patient's hospital visit. Note that we utilize the generated BHC for the subsequent generation of DI. Table 4 and 5 (Appendix A) show the output of the two target sections generated by our framework.

4.1 Section Extraction

To generate the two target sections BHC and DI, the most straightforward approach is to leverage the other readily available free-text sections in the discharge summary as the input for the generation stage. But using them all for one-stage generation is overwhelming and prone to hallucination because some sections are irrelevant or contain thousand words of nonessential information, making key aspects of the patient's record often be omitted. We thereby design sets of heuristics (e.g., regular expressions) to selectively extract clinical notes information from 13 relevant sections of the discharge summaries (e.g, *History of Present Illness*, *Pertinent Results*, . . .), with definition of each section described in Figure 1. We report data distribution of these sections in Table 6 (Appendix B)

4.2 Radiology Report Selection

Through exploring the format of different sections of the discharge summary, we notice a great complications in the structure and content of the *Pertinent Results* section, likely due to note bloat and information overload. This section, intended to highlight key findings of radiologist to the patient's treatment, is often cluttered with excessive laboratory and imaging data (e.g., blood tests, CT scans). These extraneous details can lead to challenges such as hallucination and high resource demands in generative tasks. Consequently, we explored using radiology reports as a viable alternative. These reports, often duplicated partially or entirely in the Pertinent Results section, succinctly convey diagnoses corresponding to specific lab results. We selected radiology reports with similar Impressions to those in the Pertinent Results and used these as a substitute, streamlining the content effectively.

4.3 Target Section Generation

In this framework, we performed instruction-finetuning on LLM to adapt the model to DSD. For computational feasiblity, we employed Low-Rank Adaptation (LoRA) (Hu et al., 2021), a parameter-efficient fine-tuning method that adds a small number of trainable parameters to the model while freezing the model's original weight, resulting in standalone adapters. The adapters, specifically fine-tuned for each generation task in DSD, adjust important weight of LLMs to capture and generate clinical information in the corresponding form.

Prompting Strategies Following OpenAI's prompt engineering guidelines [2], we structured our prompts into five parts, detailed in Table 7 (Appendix D): 1) Context of the discharge summary input to be summarized 2) Definition of the generation task and the specific section for documenting the discharge summary 3) Structure of the expected output of the generating section, infused with 4) Set of Chain-of-Thought (CoT) questions expected to be answered by the LLMs to capture and generate the information in each subsection of the output. Of those, our primary strategy is Part 5, which involved curating effective and generalizable CoT questions based on analysis of numerous samples. This manual effort helped in designing templates and questions that effectively guide the LLMs to focus on critical information amidst the extensive data and noise in the discharge summaries. We analyzed the medical questionnaire essential for each section, based on hundreds of samples, in Appendix C , which underpins our CoT questions and prompt design.

5 Experiments

5.1 Baseline and Implementation Details

To showcase the utility of prompt designing for adaptation to the DSD task, we developed three baselines, corresponding to three prompt variants for instruction-finetuning LLMs. $\text{Discharge\_LLM}_{\text{Base}}$ was fine-tuned with no instruction, but only the discharge summaries as input and the respective target section as output. $\text{Discharge\_LLM}_{\text{Context}}$ was fine-tuned with additional natural langauge instructions as prefix to the discharge summary to provide the context and definition of the task's input/output. Finally, $\text{Discharge\_LLM}_{\text{CoT}}$ was fine-tuned using prompts outlining the structure of the respective generating target section. Along the structure, we embed some CoT questions to elicit LLMs to generate output aligned with the questions.

We choose `Mistral` [3] (Jiang et al., 2023) as our LM. The LLM was fine-tuned on a NVIDIA RTX 4090 GPU, and took 10 hours for fine-tuning each generation task. The following hyperparameters were used: 1 sample per device, a LoRA rank and alpha of 128 and 64 for parameter-efficient fine-tuning, a learning rate of $2e10-4$. We keep other hyperparameters to their default values.

Metrics We followed the organizers to measure textual similarity and factual correctness of the generated text based on several metrics, including BLEU-4 (Papineni et al., 2002), ROUGE (Lin, 2004), BERTScore (Zhang et al., 2019), Meteor (Banerjee and Lavie, 2005), AlignScore (Zha et al., 2023), and MEDCON (Yim et al., 2023).

Dataset The dataset for this task was sourced from the MIMIC-IV (Johnson et al., 2023) dataset, including 109,168 emergency department (ED) admissions and were split into a training (68,785), a validation (14,719), a phase I testing (14,702), and a phase II testing (10,962) subsets.

Data Preprocessing To address the variation in discharge summary length, we select data within the interquartile range (Q1-Q3) for training and validation. We further ensure consistency by selecting only samples with discharge summaries containing all 13 common sections and their target sections follow the most common format, as outlined in Figure 1. Overall, 11.1k and 8.7k samples were selected for training of BHC and DI generation, respectively. For experiment, due to runtime and computational limitations, we sample 250 hidden entries from each phase's testing data, totaling 500 samples for evaluation of each generation task.

5.2 Results

Table 1 presents the performance of models fine-tuned by different prompt variants. Overall, in both generation tasks, natural language instructions plays a critical role in guiding the LLM with comprehensive knowledge to understand the task. Providing well-described context of the generation task already helps the model achieves up to 14% of

[2] https://platform.openai.com/docs/guides/prompt-engineering

[3] https://mistral.ai/

Framework	R-1	R-2	R-L	BLEU	BERTScore	Meteor	AlignScore	MEDCON
Discharge_LLM$_{\text{CoT}}$	**0.283**	0.087	0.170	**0.062**	**0.368**	**0.206**	0.230	**0.408**
Discharge_LLM$_{\text{Context}}$	0.263	**0.091**	**0.178**	0.058	0.365	0.191	**0.234**	0.397
Discharge_LLM$_{\text{Base}}$	0.240	0.074	0.159	0.043	0.347	0.170	0.221	0.376

(a) Brief Hospital Course Generation

Framework	R-1	R-2	R-L	BLEU	BERTScore	Meteor	AlignScore	MEDCON
Discharge_LLM$_{\text{CoT}}$	**0.392**	**0.151**	**0.246**	**0.077**	**0.373**	**0.272**	**0.288**	**0.452**
Discharge_LLM$_{\text{Context}}$	0.356	0.103	0.205	0.075	0.360	0.272	0.286	0.429
Discharge_LLM$_{\text{Base}}$	0.335	0.102	0.215	0.041	0.324	0.181	0.251	0.318

(b) Discharge Instructions Generation

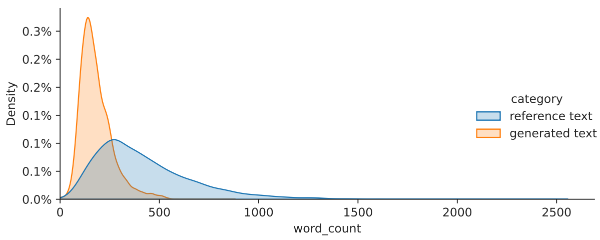
Table 1: Evaluation of prompt variants for finetuning Discharge-LLM

Framework	R-1	R-2	R-L	BLEU	BERTScore	Meteor	AlignScore	MEDCON	**Overall**
Discharge_LLM$_{\text{CoT}}$	0.370	0.131	0.245	0.068	0.360	0.314	0.215	0.324	**0.253**
Best ranked system	0.453	0.201	0.308	0.124	0.438	0.403	0.315	0.411	**0.332**

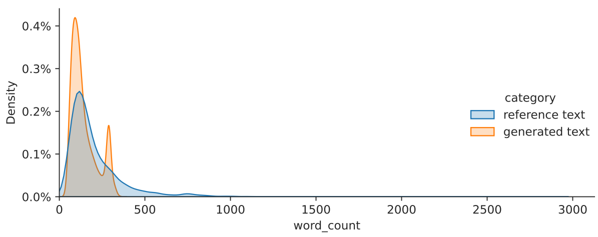
Table 2: Overall performance on two target section from the shared task's phase 2 leaderboard

Target Section	*min*	*median*	*mean*	*max*
BHC	22	367	425	2439
DI	10	153	201	2900

Table 3: Statistics of reference text's word count on phase 2's test set



(a) Brief Hospital Course



(b) Discharge Instructions

Figure 2: Distribution of samples per number of words on phase 2's test set

performance gain across the metrics and tasks. Further, infusing CoT questions into the instructions effectively elicit LLM to think better, providing an additional 9% performance increase. Notably, a reasonable improvement on MEDCON score also indicates better accuracy and consistency of clinical concepts in the generated text.

5.3 Shared Task's Evaluation Results

Table 2 summarizes our framework's overall performance on the phase 2 test set of the shared task, alongside the best-ranked system [4]. We notice there is still a gap between our Discharge_LLM$_{\text{CoT}}$ framework and the best ranked system, of which the Overall score is 0.332. This performance dip is common across submissions, likely due to prevalent data quality issues in the Discharge Summary Documentation (DSD) task. DSD, a real-world summarization challenge, involves processing actual discharge summary information with significant variability in formatting and length. Figure 2 shows word distribution variances in the target sections. It is noticeable our models are trying to set a common length for the target sections, and are struggling to converge to the wide range of lengths of the reference text, highlighted by Table 3. We note slight performance variation of Discharge_LLM$_{\text{CoT}}$ in terms of Meteor, AlignScore and MEDCON, in between Table 1 and 2. A possible reason for such variation may be that hidden test data in the shared task has distributions significantly deviate from the publicly released data for model development. Furthermore, due to computational constraints and a focus on high-quality data, we utilized only a subset of the available training data. With more comprehensive training, we anticipate improved model convergence.

6 Conclusion

In this paper, we present a LLM-based framework for Discharge Summary Documentation that adopts several prompting strategies into instruction-finetuning an LLM, which enhances structural correctness and faithfulness of clinical information in generated target sections. Using small and open-source LLMs, our work also shows the feasiblity of developing and deploying future lightweight NLP systems locally for confidential clinical tasks.

[4]https://www.codabench.org/competitions/2008/

References

Satanjeev Banerjee and Alon Lavie. 2005. METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. In *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*, pages 65–72, Ann Arbor, Michigan. Association for Computational Linguistics.

Asma Ben Abacha, Yassine Mrabet, Yuhao Zhang, Chaitanya Shivade, Curtis Langlotz, and Dina Demner-Fushman. 2021. Overview of the MEDIQA 2021 shared task on summarization in the medical domain. In *Proceedings of the 20th Workshop on Biomedical Language Processing*, pages 74–85, Online. Association for Computational Linguistics.

Yanjun Gao, Dmitriy Dligach, Timothy Miller, Dongfang Xu, Matthew MM Churpek, and Majid Afshar. 2022. Summarizing patients' problems from hospital progress notes using pre-trained sequence-to-sequence models. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 2979–2991.

Colin Grambow, Longxiang Zhang, and Thomas Schaaf. 2022. In-domain pre-training improves clinical note generation from doctor-patient conversations. In *Proceedings of the First Workshop on Natural Language Generation in Healthcare*, pages 9–22, Waterville, Maine, USA and virtual meeting. Association for Computational Linguistics.

Edward Hu, Yelong Shen, Phil Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Lu Wang, and Weizhu Chen. 2021. Lora: Low-rank adaptation of large language models. *Preprint*, arXiv:2106.09685.

Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, Lélio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. 2023. Mistral 7b. *Preprint*, arXiv:2310.06825.

Alistair EW Johnson, Lucas Bulgarelli, Lu Shen, Alvin Gayles, Ayad Shammout, Steven Horng, Tom J Pollard, Sicheng Hao, Benjamin Moody, Brian Gow, et al. 2023. Mimic-iv, a freely accessible electronic health record dataset. *Scientific data*, 10(1):1.

Andrew Lampinen, Ishita Dasgupta, Stephanie Chan, Kory Mathewson, Mh Tessler, Antonia Creswell, James McClelland, Jane Wang, and Felix Hill. 2022. Can language models learn from explanations in context? In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 537–563, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Chin-Yew Lin. 2004. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pages 74–81, Barcelona, Spain. Association for Computational Linguistics.

Zhaojiang Lin, Andrea Madotto, and Pascale Fung. 2020. Exploring versatile generative language model via parameter-efficient transfer learning. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 441–459, Online. Association for Computational Linguistics.

Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. 2023. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Comput. Surv.*, 55(9).

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics.

Dave Van Veen, Cara Van Uden, Maayane Attias, Anuj Pareek, Christian Bluethgen, Malgorzata Polacin, Wah Chiu, Jean-Benoit Delbrouck, Juan Zambrano Chaves, Curtis Langlotz, Akshay Chaudhari, and John Pauly. 2023. RadAdapt: Radiology report summarization via lightweight domain adaptation of large language models. In *The 22nd Workshop on Biomedical Natural Language Processing and BioNLP Shared Tasks*, pages 449–460, Toronto, Canada. Association for Computational Linguistics.

Justin Xu, Zhihong Chen, Andrew Johnston, Louis Blankemeier, Maya Varma, Jason Hom, William J. Collins, Ankit Modi, Robert Lloyd, Benjamin Hopkins, Curtis Langlotz, and Jean-Benoit Delbrouck. 2024. Overview of the first shared task on clinical text generation: Rrg24 and "discharge me!". In *The 23rd Workshop on Biomedical Natural Language Processing and BioNLP Shared Tasks*, Bangkok, Thailand. Association for Computational Linguistics.

Wen-wai Yim, Yujuan Fu, Asma Ben Abacha, Neal Snider, Thomas Lin, and Meliha Yetisgen. 2023. Aci-bench: a novel ambient clinical intelligence dataset for benchmarking automatic visit note generation. *Scientific Data*, 10(1):586.

Yuheng Zha, Yichi Yang, Ruichen Li, and Zhiting Hu. 2023. AlignScore: Evaluating factual consistency with a unified alignment function. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 11328–11348, Toronto, Canada. Association for Computational Linguistics.

Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q Weinberger, and Yoav Artzi. 2019. Bertscore: Evaluating text generation with bert. *arXiv preprint arXiv:1904.09675*.

A Generated Output of Brief Hospital Course and Discharge Instructions

This section presents details of Table 4 and 5, which show the output of the two "Brief Hospital Course" and "Discharge Instructions" target sections generated by our framework, taken from medical information of patient with $hadm\_id = 21720538$ from the phase 2's test set.

B Data Distribution of Discharge Summary Sections

Table 6 presents the percentage distribution of common sections in the discharge summary text of the training, validation and testing subsets.

C Questionnaire for Discharge Summary Documentation

C.1 Brief Hospital Course

- *Patient Background and Presenting Complaint*: "What is the patient's background including pre-existing medical conditions, and what symptoms or events led to their current hospital admission?"

- Key Diagnoses and Evaluations: "What are the key diagnoses identified during the hospital stay? For each, how was the diagnosis reached, including any significant tests or evaluations conducted?"

- Treatment and Management Strategies: "What were the main treatment strategies employed for the patient's conditions during their stay? Include medications adjusted, procedures performed, and any therapeutic interventions."

- Complications and Additional Diagnoses: "Were there any complications or additional diagnoses during the hospital stay? How were these addressed and managed?"

- Progress and Monitoring: "How did the patient's condition progress throughout the hospital stay, including any monitoring of symptoms, response to treatments, and adjustments made to the treatment plan?"

- Support and Consultation Services: "Which specialist services or support consultations were involved in the patient's care? How did these consultations impact the patient's treatment plan and recovery?"

- Discharge Planning and Instructions: "What were the conditions and considerations for the patient's discharge? Include the discharge medications, any changes from previous medication regimens, and follow-up care or lifestyle recommendations."

- Follow-Up and Post-Discharge Care: "What are the specific follow-up care instructions and any scheduled tests or consultations? Highlight the importance of follow-up for managing ongoing conditions or monitoring recovery."

C.2 Discharge Instructions

- Initial Assessment and Diagnosis: What led to the patient's admission to the hospital, and what were the initial symptoms? Based on the patient's symptoms, what diagnoses were considered and which was confirmed?

- Treatment and Hospital Stay: What treatments were provided to address the patient's symptoms or condition during the hospital stay? Were any surgeries recommended or performed? If a surgery was recommended but not performed, what were the reasons? What were the outcomes of the treatments or interventions provided?

- Patient's Decisions and Care Preferences: Did the patient make any specific requests regarding their care, such as refusing a treatment or requesting a transfer? How were these handled? How did the patient's decisions affect their treatment plan and discharge process?

- Comprehensive Post-Discharge Instructions: What are the general care instructions for the patient after discharge, including diet, activity level, and medication management? Are there any specific symptoms or signs that the patient should monitor for which would require immediate medical attention? How should the patient manage their regular home medications in addition to any new medications prescribed at discharge?

- Activity and Lifestyle Recommendations: What specific activity restrictions or recommendations are given to ensure a smooth recovery? (e.g., weight lifting limits, mobility advice) Are there any restrictions on driving or

```
___ with PMH of NASH vs cryptogenic cirrhosis, HTN,
hyperlipidemia, hypothyroidism who presents from ___
___ with DVT and renal failure.

# DVT: Patient with extensive RLE DVT extending from the right
calf veins into the popliteal vein and femoral vein. Patient
was started on heparin gtt. He was evaluated by ___ who
recommended discharge home with home ___.

# Renal Failure: Patient presented with Cr 8.9, likely
pre-renal in etiology. Patient was given albumin and
diuresis. Cr improved to 3.5 on discharge.

# Cirrhosis: Patient with cirrhosis secondary to NASH vs
cryptogenic cirrhosis. Patient was continued on lactulose and
rifaximin. He was also continued on nadolol.

# Hepatic Encephalopathy: Patient with history of hepatic
encephalopathy. Patient was continued on lactulose and
rifaximin.

# Ascites: Patient with large volume ascites. Patient was
continued on diuresis.

# HTN: Patient was continued on home nadolol.

# Hypothyroidism: Patient was continued on home levothyroxine.

# GERD: Patient was continued on home omeprazole.

# Osteopenia: Patient was continued on home calcium and vitamin
D.

# Allergic Rhinitis: Patient was continued on home cetirizine.

# Chronic Kidney Disease: Patient with chronic kidney disease
secondary to cirrhosis. Patient was continued on home
albumin/dietary management.

# Code: DNR/DNI
# Emergency Contact: ___ (daughter) ___

TRANSITIONAL ISSUES:
- Patient will need to follow up with hepatology and ___
clinic.
- Patient will need to follow up with ___ clinic for
continued management of his DVT.
- Patient will need to follow up with ___ clinic for
continued management of his ascites.
- Patient will need to follow up with ___ clinic for
continued management of his hepatic encephalopathy.
- Patient will need to follow up with ___ clinic for
continued management of his cirrhosis.
- Patient will need to follow up with ___ clinic for
continued management of his hypertension.
- Patient will need to follow up with ___ clinic for
continued management of his hyperlipidemia.
```

Table 4: Generated Output of Brief Hospital Course.

operating machinery, especially if the patient is taking new or continued pain medication?

- Follow-up Care and Monitoring: What follow-up appointments or tests are recommended for the patient? With whom should these appointments be made? How should the patient approach symptom management, especially if they experience pain, dehydration, or other concerning symptoms?

- Communication with Healthcare Providers:

```
Dear Mr. ___,

It was a pleasure taking care of you at ___
___. You were admitted to the hospital because you
were having kidney failure. You were treated with medications
to help your kidneys recover. You also had a blood clot in your
legs. You were started on blood thinners to help prevent
further clots. You were also found to have a lot of fluid in your
belly. You were given medications to help remove the fluid.

You were also found to have a lot of confusion. You were given
medications to help with this. You were also found to have a
blood infection. You were treated with antibiotics. You were
discharged home with hospice care.

We wish you the best.

Sincerely,
Your ___ Team
```

Table 5: Generated Output of Discharge Instructions.

Section	**train**	**valid**	**test (phase 1)**	**test (phase 2)**
Allergies	0.999941669	0.999795571	0.999864232	0.999453801
Chief Complaint	0.999956252	0.999863714	1	0.9997269
Major Surgical or Invasive Procedure	0.518607636	0.516456559	0.517615912	0.517979062
History of Present Illness	0.980386152	0.982010221	0.980788813	0.97997269
Past Medical History	0.960203576	0.96197615	0.958387075	0.960491579
Social History	0.97414472	0.976149915	0.974204059	0.973782431
Family History	0.967567883	0.968245315	0.968094495	0.966317706
Physical Exam	0.978315397	0.980102215	0.978141335	0.977878926
Pertinent Results	0.981231954	0.98153322	0.981942842	0.981429222
Brief Hospital Course	1	1	1	1
Medications on Admission	0.939787675	0.939625213	0.937750322	0.939007738
Discharge Medications	0.980590311	0.981192504	0.980585161	0.981975421
Discharge Disposition	0.989121241	0.987189097	0.987984522	0.987528448
Discharge Diagnosis	0.991950302	0.992231687	0.992261218	0.993172508
Discharge Condition	0.999970834	0.999931857	1	1
Discharge Instructions	1	1	1	1

Table 6: Data Distribution of sections in the discharge summaries in the provided dataset

Under what circumstances should the patient immediately contact their healthcare provider or seek emergency care? What is the recommended way for the patient to communicate with their healthcare team (e.g., phone call, hospital return)?

- Encouragement and Support: How can we encourage the patient to adhere to their discharge instructions and reassure them about their recovery process? What resources or support systems can we recommend to the patient for additional help or information post-discharge?

D Prompts for Discharge Summary Documentation

We present the prompts for generation of the two critical “Brief Hospital Course” and “Discharge Instructions” target sections in Table 7

Prompt for Brief Hospital Course Generation	Prompt for Discharge Instructions Generation
In this task, you are provided with a Discharge Summary delimited by triple quotes. Discharge Summaries are documents that outline the care a patient received during their hospital stay, including diagnoses, treatments, and follow−up care instructions, prepared at the time of a patient's discharge. Discharge Summaries are split into various sections and written under a variety of headings, relating to admission, diagnosis and relevant discharge information. But the provided Discharge summary will be missing the \"Brief Hospital Course\". \"Brief Hospital Course\" is a section of the discharge summaries that outlines the key events of a patient's hospital stay, including the progression from admission to discharge. It is written for the subsequent care providers about the critical aspects of the patient. You are tasked to generate the missing \"Brief Hospital Course\" section in the discharge summary, based on the information of other sections in the discharge summary. Brief Hospital Course outlines the key events of a patient's hospital stay, including the progression from admission to discharge. It is written for the subsequent care providers about the critical aspects of the patient The summary should be written in the following structure, by answering some important questions: 1. Initial presentation: Describe the patient's initial presentation, including the main complaint and relevant history. ∗ What were the main treatment strategies employed for the patient's conditions during their stay? Include medications adjusted, procedures performed, and any therapeutic interventions. ∗ What are the key diagnoses identified during the hospital stay? 2. Treatment course: – For each section header named by "#Condition Name", provide a detailed description of each condition, disease, or symptom of the patient by answering the following questions: ∗ What is the patient's background relating to the condition, disease, or symptom ∗ Describe the treatment strategy, including any medications given, procedures performed, and dietary adjustments. ∗ How was the diagnosis reached, including any significant tests or evaluations conducted? ∗ What were the significant medical or surgical interventions during the hospital stay, including any procedures, diagnostic tests (e.g., CT Scan, Imaging, Blood Test, MRI), and changes in medication? ∗ Were there any complications or additional diagnoses during the hospital stay? How were these addressed and managed? ∗ How did the patient's condition progress throughout the hospital stay, including any monitoring of symptoms, response to treatments, and adjustments made to the treatment plan? ∗ What were the conditions and considerations for the patient discharge? Include the discharge medications, any changes from previous medication regimens, and follow−up care or lifestyle recommendations. 3. Transitional issues: Highlight any transitional care issues addressed during the hospital stay, including changes in medication, dietary adjustments, and specific care instructions. 4. Acute/active issues: Detail the management of acute or active issues encountered during the stay, using the provided structure for each condition. 5. Chronic/stable issues: Summarize how chronic conditions were managed during the stay and any adjustments made to long−term management plans.	In this task, you are provided with a Discharge Summary delimited by triple quotes. Discharge Summaries are documents that outline the care a patient received during their hospital stay, including diagnoses, treatments, and follow−up care instructions, prepared at the time of a patient's discharge. Discharge Summaries are split into various sections and written under a variety of headings, relating to admission, diagnosis and relevant discharge information. But the provided Discharge summary will be missing the \"Discharge Instructions\". \"Discharge Instructions\" is a section of the discharge summaries that summarizes key events of a patient's hospital stay, including the progression from admission to discharge. and provide detailed guidelines to patients (and often their caregivers) upon discharge from a hospital or healthcare facility, outlining how to care for themselves at home. You are tasked to generate the missing \"Discharge Instructions\" section in the discharge summary, based on the information of other sections in the discharge summary. Discharge Instructions summarizes key events of a patient's hospital stay, including the progression from admission to discharge. and provide detailed guidelines to patients (and often their caregivers) upon discharge from a hospital or healthcare facility, outlining how to care for themselves at home. The summary should be written in the following structure, by answering some important questions: 1. Admission Reason: A concise explanation of: ∗ Why the patient was admitted, including any specific conditions or symptoms addressed during the stay. ∗ What was the patient's diagnosis upon admission to the hospital? 2. Hospital Course: A concise summary of what happened to the patient's at the hospital ∗ How was the diagnosis reached, including any significant tests or evaluations conducted (e.g., CT Scan, Imaging, Blood Test, MRI)? ∗ What were the significant medical or surgical interventions during the hospital stay , including any procedures, diagnostic tests (e.g., CT Scan, Imaging, Blood Test, MRI), and changes in medication? ∗ Describe the treatment strategy, including any medications given, procedures performed, and dietary adjustments. ∗ How did the patient respond to the treatment and procedures? Did the patient make any specific requests regarding their care, such as refusing a treatment or requesting a transfer? How were these handled? ∗ Were there any complications or notable improvements in the patient's condition during the stay? ∗ What were the outcomes of the treatments or interventions provided? 3. Post−Discharge Instructions: + Follow−Up Care: ∗ What the patient should do after leaving the hospital? ∗ What specific activity restrictions or recommendations are given to ensure a smooth recovery? (e.g., weight lifting limits, mobility advice) ∗ Are there any restrictions on driving or operating machinery, especially if the patient is taking new or continued pain medication? ∗ Instructions on how to continue treatments started in the hospital, such as new medications or therapy. + Medications (Optional): ∗ Comprehensive instructions for all prescribed medications, including dosage, timing, and any specific instructions for use. ∗ How should the patient manage their regular home medications in addition to any new medications prescribed at discharge? + Monitoring: ∗ Guidelines on any self−monitoring the patient should perform at home, such as weighing themselves, monitoring blood pressure, or blood sugar levels, with instructions on when to contact their healthcare provider. ∗ Are there any specific symptoms or signs that the patient should monitor for which would require immediate medical attention? Under what circumstances should the patient immediately contact their healthcare provider or seek emergency care?

Table 7: Prompts for “Brief Hospital Course” and “Discharge Instructions” generation of the Discharge-LLM framework.