What's Going On with Me and How Can I Better Manage My Health? The Potential of GPT-4 to Transform Discharge Letters into Patient-Centered Letters to Enhance Patient Safety: A prospective, exploratory study

Felix Eisinger, Friederike Holderried, Moritz Mahling, Christian Stegemann–Philipps, Anne Herrmann–Werner, Eric Nazarenus, Alessandra Sonanini, Martina Guthoff, Carsten Eickhoff, Martin Holderried

Submitted to: Journal of Medical Internet Research
on: October 04, 2024

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on it's website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressively prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

What's Going On with Me and How Can I Better Manage My Health? The Potential of GPT-4 to Transform Discharge Letters into Patient-Centered Letters to Enhance Patient Safety: A prospective, exploratory study

Felix Eisinger[1] MD; Friederike Holderried[2] MD; Moritz Mahling[3] MD; Christian Stegemann–Philipps[2]; Anne Herrmann–Werner[2] MD; Eric Nazarenus[2]; Alessandra Sonanini[2] MD; Martina Guthoff[1] MD; Carsten Eickhoff[4] BS, MS, PhD; Martin Holderried[3] MD

[1]Department of Diabetology, Endocrinology, Nephrology University of Tübingen Tübingen DE
[2]Tübingen Institute for Medical Education (TIME) University of Tübingen Tübingen DE
[3]Department of Medical Strategy, Process and Quality Management University Hospital Tübingen Tübingen DE
[4]Institute for Bioinformatics and Medical Informatics University of Tübingen Tübingen DE

Corresponding Author:
Friederike Holderried MD
Tübingen Institute for Medical Education (TIME)
University of Tübingen
Elfriede-Aulhorn-Str. 10
Tübingen
DE

Abstract

Background: For hospitalized patients, the discharge letter is an important source of medical information, containing numerous discharge instructions and health care tasks for the patients to manage their own health. However, it is usually written in professional jargon that is inaccessible to patients with little medical knowledge. Large language models such as GPT have the potential to translate discharge summaries into patient-friendly letters.

Objective: In this study, we used GPT-4 to transform discharge letters into more readable patient letters and evaluated how comprehensively patient safety-relevant information was identified and transferred from the discharge letters into patient-centered letters.

Methods: We developed three discharge letters based on common medical conditions with 72 patient safety-relevant information, defined as “learning objectives.” Then, we prompted GPT-4 to transform the discharge letters into patient-centered letters. The patient letters were analyzed for medical quality, patient-centricity, and the potential to identify and translate the learning objectives. Bloom’s taxonomy was used to analyze and categorize learning objectives.

Results: While GPT-4 addressed the majority (56/72; 78%) of the learning objectives from the discharge letters, 11 of the 72 learning objectives (15%) were not included in the majority of the patient-centered letters. A qualitative analysis based on Bloom’s taxonomy showed that learning objectives of the Bloom category Understand (9/11) were more frequently missed than those of the Bloom category Remember (2/11). Most of the missing learning objectives pertained to the content field “prevention of complications,” while learning objectives regarding “lifestyle” and “organizational” aspects were mentioned more often. Medical errors occurred in a few (31/787; 4%) of the sentences. Regarding patient-centricity, the patient-centered letters showed better readability than the discharge letters, using fewer medical terms (132/860; 15%) than the discharge letters (165/273; 60%), as well as fewer abbreviations (43/860; 5%) versus (49/273; 18%) and more explanations of medical terms (121/131; 92%) versus (0/165; 0%).

Conclusions: Conclusion:
Our study shows that GPT-4 has the potential to transform discharge letters into more patient-centered letters. However, while readability and patient-centricity are already well-established, the patient-centered letters do not comprehensively address all patient safety-relevant information, leading to the omission of some important aspects. Further optimization in the prompt-engineering might help to overcome this issue.

Conclusion: Our study shows that GPT-4 has the potential to transform discharge letters into more patient-centered letters.

However, while readability and patient-centricity are already well-established, the patient-centered letters do not comprehensively address all patient safety-relevant information, leading to the omission of some important aspects. Further optimization in the prompt-engineering might help to overcome this issue.

(JMIR Preprints 04/10/2024:67143)
DOI: https://doi.org/10.2196/preprints.67143

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

What's Going On with Me and How Can I Better Manage My Health? The Potential of GPT-4 to Transform Discharge Letters into Patient-Centered Letters to Enhance Patient Safety: A Prospective, Exploratory Study

Felix Eisinger[1,2,3], Friederike Holderried[4], Moritz Mahling[6], Christian Stegemann–Philipps[4], Anne Herrmann–Werner[4], Eric Nazarenus[4], Alessandra Sonanini[4,5], Martina Guthoff MD **[1,2,3]**, Carsten Eickhoff[7], Martin Holderried[6]

[1]Department of Diabetology, Endocrinology, Nephrology, University of Tübingen, Tübingen, Germany

[2] Institute for Diabetes Research and Metabolic Diseases of the Helmholtz Center Munich at the University of Tübingen, Tübingen, Germany.

[3] German Center for Diabetes Research (DZD e.V.), Neuherberg, Germany.

[4] Tübingen Institute for Medical Education (TIME), Eberhard Karls University, Tübingen, Germany

[5] Department of Gastroenterology, Hepatology and Infectiology, University Hospital Tübingen, Tübingen, Germany

[6] Department of Medical Strategy, Process and Quality Management, University Hospital Tübingen, Tübingen, Germany

[7] Institute for Bioinformatics and Medical Informatics, University of Tübingen, Tübingen

*Corresponding Author

Friederike Holderried, MD
Tübingen Institute for Medical Education (TIME), Eberhard Karls University, Tübingen

E-Mail: Friederike.Holderried@med.uni-tuebingen.de

Abstract

Background:
For hospitalized patients, the discharge letter is an important source of medical information, containing numerous discharge instructions and health care tasks for the patients to manage their own health. However, it is usually written in professional jargon that is inaccessible to patients with little medical knowledge. Large language models such as GPT have the potential to translate discharge summaries into patient-friendly letters.

Objective:
In this study, we used GPT-4 to transform discharge letters into more readable patient letters and evaluated how comprehensively patient safety-relevant information was identified and transferred from the discharge letters into patient-centered letters.

Methods:
We developed three discharge letters based on common medical conditions with 72 patient safety-relevant information, defined as “learning objectives.” Then, we prompted GPT-4 to transform the discharge letters into patient-centered letters. The patient letters were analyzed for medical quality, patient-centricity, and the potential to identify and translate the learning objectives. Bloom’s taxonomy was used to analyze and categorize learning objectives.

Results:
While GPT-4 addressed the majority (56/72; 78%) of the learning objectives from the discharge letters, 11 of the 72 learning objectives (15%) were not included in the majority of the patient-centered letters. A qualitative analysis based on Bloom’s taxonomy showed that learning objectives of the Bloom category Understand (9/11) were more frequently missed than those of the Bloom category Remember (2/11). Most of the missing learning objectives pertained to the content field “prevention of complications,” while learning objectives regarding “lifestyle” and “organizational” aspects were mentioned more often. Medical errors occurred in a few (31/787; 4%) of the sentences. Regarding patient-centricity, the patient-centered letters showed better readability than the discharge letters, using fewer medical terms (132/860; 15%) than the discharge letters (165/273; 60%), as well as fewer abbreviations (43/860; 5%) versus (49/273; 18%) and more explanations of medical terms (121/131; 92%) versus (0/165; 0%).

Conclusion:
Our study shows that GPT-4 has the potential to transform discharge letters into more patient-centered letters. However, while readability and patient-centricity are already well-established, the patient-centered letters do not comprehensively address all patient safety-relevant information, leading to the omission of some important aspects. Further optimization in the prompt-engineering might help to overcome this issue.

Keywords: GPT-4, patient letters, health care communication, artificial intelligence, patient safety, patient education

Introduction

Ensuring patient safety is fundamental in healthcare. A key aspect of patient safety is adherence to treatments and interventions prescribed by the medical provider, as it is essential for the prevention of long-term disease progression, reduction of complications and improvement in quality of life [1]. In clinical practice, however, non-adherence is a widespread problem. According to the WHO, only 50% of patients with chronic diseases in developed countries adhere to their prescribed therapy regimens [2]. A common factor driving non-adherence is patients` lack of understanding of their disease and the underlying principles of therapy [3]. Historically, the patient–healthcare worker relationship has followed a “paternalistic” model, in which the patient has been a “passive spectator on his or her own healing process” [4]. Fortunately, this dynamic has changed in recent years. However, empowering patients to comprehensively understand their individual health issues remains a promising approach to improving adherence and thus promoting patient safety [3].

For hospitalized patients, the discharge letter is an important source of medical information that complements the conversation with the physician. It also plays an important role in ensuring communication between hospital doctors and other healthcare providers, such as primary care physicians [5]. Good communication at transition from inpatient to outpatient care is crucial for patient safety [6]. Forster et al. found that 59% of preventable adverse events after hospital discharge were due to poor communication between the hospital caregivers and either the patient or the primary care physician [7]. After discharge, particularly patients with chronic conditions face numerous self-management challenges, such as adhering to prescribed medication regimens, maintaining a specific diet, and engaging in physical activity. Any failure to adhere to these aspects can have serious health consequences.

However, in clinical routine, discharge letters are typically addressed to the general practitioners or other medical professionals, and are thus laden with professional jargon that is inaccessible to patients with limited medical knowledge [8]. The development of patient-centered discharge letters with improved readability for the patients has been shown to enhance patient understanding [9, 10]. In the hospital setting, however, there is often limited time for the preparation of additional, individualized patient letters, as a significant portion of working hours is spent on non-patient-related tasks and documentation [11].

In this context, advances of artificial intelligence (AI) offer a promising approach to providing personalized and scalable support for helping patients to understand medical information. Various studies have shown the substantial medical knowledge of large language models (LLMs), such as a generative pretrained transformer (GPT)[12-15]. Zaretsky et al. has utilized LLMs to translate discharge summaries into patient-friendly language, addressing common readability metrics and the Patient Education Materials Assessment Tool (PEMAT) scoring. However, they encountered significant limitations in both accuracy and completeness [16]. On the other hand, the use of readability scores has been controversially discussed in the literature [17]. In our own work, we have demonstrated GPT-4’s ability to answer psychosomatic medicine exam questions [18]. A qualitative analysis of incorrect answers, based on Bloom’s revised taxonomy [19, 20] showed that errors varied depending on the cognitive level. It remains unclear whether this effect can also be observed in patient letters developed by GPT-4. Furthermore, the extent to which GPT-4 addresses comprehensive patient safety-relevant information—a key aspect of the discharge letter—has yet to be clarified.

In the present study, we further investigated another topic. We used GPT-4 to transform discharge letters into accessible patient-centered letters and evaluated its ability to identify and incorporate patient safety-relevant information. To pinpoint potential errors or gaps in the AI-driven transformation process, patient safety-relevant information was categorized and analyzed according to Bloom’s revised taxonomy (Remember, Understand) [19, 20].

In summary, our study addresses the following questions.

Major objective:

How comprehensively does GPT-4 identify and transform patient safety-relevant information, as measured by the learning objectives, from discharge letters into patient-centered letters?

Minor objectives:

A) How do GPT-4 generated patient-centered letters perform in terms of medical correctness (measured by medical accuracy, case-specific relevance and the sources of information used in the constituent sentences)?

B) How well do GPT-4-generated patient-centered letters perform in terms of patient-centricity compared to discharge letters (with patient-centered language measured by standard readability scores, word and sentence count, and the use of medical jargon, explanations, abbreviations, repetitions, and direct addressing)?

Material and Methods

1. Study Outline

Since the primary goal of patient-centered letters is to convey important information to patients, aligned with specific didactic 'learning objectives', we defined learning objectives for three common medical conditions, including the corresponding competence level according to Bloom`s taxonomy. Based on these learning objectives and the associated medical conditions, we created three discharge letters (**Figure 1, Supplementary Material 1**). We then prompted GPT-4 to generate a patient-centered letter from each discharge letter (**Supplementary Material 1)**. Considering the variability in GPT-4`s output, we repeated the generation of the patient-centered letter five times per discharge letter with the same prompt.

The resulting patient-centered letters were analyzed by a team of two experienced clinicians (FE, FH) in terms of medical quality, patient-centricity, and their potential to convey safety-relevant medical information.

2. Development of Discharge Letters

For development of the discharge letters, we meticulously structured a multistep process (**Supplemental Figure 1**) to cluster critical patient information, which GPT-4 was prompted to extract from the discharge letters at various structural levels. These levels enabled us to assess GPT-4`s effectiveness in retrieving critical patient information during the evaluation process based on multiple criteria. This approach allowed for a clearer depiction of GPT-4`s competence in this specific task area. To demonstrate the transferability of our findings, it was essential to establish the extent to which GPT-4's ability to identify patient-important information was context-independent. Accordingly, we developed three distinct scenarios, each with different disease profiles and care settings.

Stepwise approach:

- Step 1: We selected three common diseases based on their high prevalence: arterial hypertension (AHT), type 2 diabetes mellitus (DM), and diabetic kidney disease (DKD).
- Step 2: The setting for each scenario was chosen, encompassing either outpatient consultations with diagnostics or inpatient treatment for initial diagnosis or follow-up care.
- Step 3: To construct relevant educational content for each scenario, we aligned all learning content with the cognitive process dimension of factual knowledge in Bloom’s revised taxonomy [20].
- Step 4: The educational content was then categorized into four distinct content fields: (1) organizational aspects, (2) medication protocols, (3) prevention of complications, and (4) disease management and lifestyle changes.
- Step 5: For each piece of learning content, we developed two corresponding learning objectives (**Supplementary Material 1**): one at the Remember level (e.g., the patient knows that he/she has to conduct a 24-hour urine collection test), and another at the Understand level (e.g., the patient understands that the 24-hour urine collection is necessary to rule out hormonal causes of hypertension). Each discharge letter thus comprised 12 pieces of learning content, distributed across the four content fields, resulting in 24 learning objectives per letter.
- Step 6: These learning objectives were then systematically integrated into the format of a typical discharge letter. The discharge letters were developed by two experienced physicians (FE, FH) and validated for face validity by two additional board-certified and experienced internal medicine and nephrology specialists (MM, MG).

3. Prompt Development and Creation of Patient-centered Letters

To the best of our knowledge, there is currently no ideal model for patient-centered letter structure. Our group has focused, first, on expert-formulated learning objectives based on the content deemed important from a medical perspective and, second, on patient-centered language. This provided us with a basic letter structure that we could aim for.

We utilized GPT-4 (gpt-4-0613, accessed December 2023) for generating the patient-centered letters, maintaining the model`s default parameters and setting the temperature to the default value of 1.0. The AI-generated patient-centered letter development process was divided into multiple stages (**Supplemental Figure 2**), leveraging methodologies involving large language models (LLMs) as agents.

Stage 1: Essential descriptions from the original discharge letters were transformed into concise, comprehensible summaries, termed "structured info," using a specialized system prompt for data extraction.

Stage 2: A general summary of the original letter was created to provide an overview of the information, utilizing a straightforward system prompt without specific constraints to minimize redundancy.

Stage 3: This final stage focused on extracting detailed "action points" critical for patient understanding and compliance. This multistep process, known as ‘prompt chaining,’ began with segmenting the original letter to isolate significant instructions and reasons for necessary behavioral changes. Each segment was processed to reduce redundancy and enhance clarity, using structured prompts to request information in a predefined JSON format.

Further refinement involved adversarial prompting to identify and simplify complex medical terms and jargon, tailoring the content to a comprehension level equivalent to an eighth-grade reading ability [21]. The culmination of this process was a restructured and more comprehensible summary of actionable points, intended to effectively communicate essential medical information to patients.

4. Analysis of Patient-centered and Discharge Letters

4.1 Rating Process

Two experienced clinicians (FE, FH) independently rated the discharge letters and the patient-centered letters. For the rating process, the letter was divided into sentences. A sentence was defined as a unit of one or more words ending with a period, a colon or a new paragraph (typically, but not exclusively, representing complete sentences). Specific titles of the patient-centered letters were predefined using established terms (e.g., the 'Main Diagnosis') and these predetermined titles were excluded from the rating.

Each sentence was assessed individually by both raters. To ensure a standardized rating process, a general rating structure was defined in advance, using an ordinal scale between 0 (not fulfilled) and 2 (fully fulfilled) (**Table 1** and **Supplementary Table 1**). In cases of uncertainty, the raters discussed the issues until consensus was reached. Results are presented as overall (patient-centered letters vs. discharge letters) and as subgroup analysis (subgroups: letters on arterial hypertension (AHT), diabetes mellitus (DM) and diabetic kidney disease (DKD)).

Table 1: Schematic Rating Scale (Learning Objectives)

Illustration of the rating scale used to analyze the learning objectives. It includes rating examples of learning objectives that were fully (2) or partially (1) rendered in the patient-centered letters.

Content field	Common disease	Bloom category	Learning objective	2 full information Rating Examples 1 partly mentioned 0 not mentioned
Medication	Type 2 diabetes mellitus	Remember	Take new medication atorvastatin 20mg in the morning	2: “One tablet (20mg) atorvastatin in the morning (…)” 1:"Atorvastatin: Take one tablet every morning.”
		Understand	Atorvastatin helps to control lipid levels.	2: “Atorvastatin: (…) can help keep cholesterol low and protect your heart.” 1: "You have also been given medication to lower high cholesterol, which is important in reducing the risk of heart disease."
Prevention of complications	Diabetic nephropathy	Remember	Avoid NSAIDs such as Ibuprofen	2:"Ibuprofen: You should avoid this painkiller (…).” 1:"It is important for you to avoid anti-inflammatory pain medications."
		Understand	Avoiding NSAIDs is important to prevent kidney damage	2: "Ibuprofen: (…) as it can impair your kidneys." 1: "You had been taking painkillers regularly, which may have also strained your kidneys."

Supplementary Table 1: Rating Scale (Medical Quality and Patient-Centricity)

This outline describes the rating scale used to evaluate the patient-centered letters for medical quality and patient-centricity. Examples from the patient-centered letters or the discharge letters are presented in italics.

Medical correctness	Definition (*and example from patient letter or discharge letter*)
Medically correct	The sentence is medically correct. *“It is important that you reduce your alcohol consumption, as alcohol can affect blood sugar levels and increase the risk of other health problems.”*
Medically incorrect	The sentence is medically not correct (normal HbA1c is below 5.7%). *“The HbA1c value, which tells me how my blood sugar has been over the last few weeks, is 14.1%, which is also very high (normal*

	below 6%).”
Case-Specific relevance	
Very relevant	Information is related to the primary diagnosis. *“During your stay in hospital, you were diagnosed with type 2 diabetes for the first time, which means that your blood glucose levels are too high.”*
Rather relevant	Information is related to the secondary diagnosis. *“You had an appendectomy in 1972.”*
Neither/nor relevant	Not attributable to a primary or secondary diagnosis. *“In plain language, your hospital stay can be summarized as follows:”*
Source of information	
Discharge letter	Information is derived from discharge letter. *“You have chronic kidney disease caused by diabetes and high blood pressure. “*
Not from discharge letter	Information was added by GPT-4. *“They performed a kidney biopsy in which they obtained a small sample of tissue from your kidney to examine it under a microscope.”*
Medical terms	
Use of medical term	Medical term is used in the sentence. *“We ask for the completion of a 24h urine collection for metanephrine to investigate a pheochromocytoma.”*
No special term	No special term is used in the sentence. *“You should also try to reduce your weight, eat more healthily (lots of fruit and vegetables, less fatty dairy products) and use less salt.”*
Explanations	
Medical term explained	Medical term is explained in everyday language. *“Sleep apnea syndrome is a condition in which breathing stops during sleep, which can also affect blood pressure.”*
Medical term not explained	Medical term is not explained. *“In addition, sleep apnea syndrome should be ruled out in the outpatient setting.”*
Abbreviations	
Use of abbreviations	An abbreviation is used in the sentence. *“CVRF:”*
No use of abbreviations	An abbreviation is not used in the sentence. *“This is called hypertension.”*
Repetitions	
Use of repetitions	Repetition of a fact in the sentence. *“It is also recommended that you regularly measure your blood pressure at home (…) Measure your blood pressure regularly at home (…)”*
No use of repetitions	No repetition of a fact in the sentence. *“Drink about 2 liters of water a day unless your doctor says otherwise.”*
Addressing	
Directly addressed	Patient is directly addressed.

	“You should take one tablet every morning.”
Not directly addressed	The patient is not directly addressed. *“We therefore recommend initiation of atorvastatin as indicated above.”*
Not applicable	Not applicable since there is no person to be addressed in the sentence. *“An ultrasound of the pancreas showed no evidence of a malignancy.”*

4.2 Learning Objectives

The GPT-4 generated patient-centered letters were rated on how many of the learning objectives were identified and addressed. To better understand GPT-4`s capability to highlight patient safety-relevant information, a descriptive analysis of the learning objectives was performed. This analysis determined whether the learning objectives were fully addressed, partially addressed, or missed entirely (**Table 1**).

4.3 Indicators of Patient-Centricity

All letters were rated for their readability using the Flesch Reading Ease test and the Swedish LIX and RIX readability indices. The Flesch Reading Ease ranges from 0 (very difficult to read) to 100 (very easy to read) [22]. The Swedish Läsbarhetsindex (LIX) typically ranges from 20 (very easy to read) to 60 (very difficult to read) [23]. The Readability Index (RIX) is a modified version of the LIX, a higher number indicates a more complex text in this score [24]. Due to the fragmented nature of some discharge letter sentences (**Supplementary Material 1**), only the sections on procedure and the medical discharge summary were used to calculate the readability score. In contrast, patient-centered letters, which were written in complete sentences, were evaluated as a whole for readability. To identify potential barriers for non-medical readers, the use of medical terms, explanations of medical terms, abbreviations, and repetitions was analyzed for each sentence (**Supplementary Table 1**). Additionally, it was examined whether the recipient of the letter (patient-centered letter: patient; discharge letter: general practitioner) was directly addressed in the respective sentence (if applicable).

4.3 Medical Correctness, Case-Specific Relevance and Source of Information

All sentences were rated for medical correctness (**Supplementary Table 1**). To gain deeper insights into medical errors in the patient letters, a qualitative analysis of the incorrect sentences was conducted using the Braun-Clarke inductive approach [25]. Sentences were also evaluated for their case-specific relevance: sentences referring to the primary diagnosis were rated as very relevant, those related to secondary diagnoses were considered rather relevant, and all other sentences were classified as neither relevant nor irrelevant, as no irrelevant information was found in the letters. Additionally, it was assessed whether the medical information in the patient letters was derived from the discharge letter or generated by GPT-4.

5. Statistical Analysis

Statistical analysis and figure generation were performed using R statistical software (version 4.3.1; R Foundation for Statistical Computing). Data are presented as total (n) and percentage or, if not normally distributed, as medians with interquartile ranges (25% and 75% quartiles). Decimal numbers were rounded to whole numbers for clarity.

Results

1. Quantitative Analysis of Letter Structure

For the discharge letters, the number of sentences per letter and the median word count per sentence showed minimal variation across the three different disease entities (**Table 2**).

For the patient letters, GPT-4 generated a total of 952 sentences, with 860 sentences remaining after excluding those predefined in the prompts. The median word count per sentence was 15 words (IQR 8–20). The number of sentences per patient-centered letter varied slightly between the medical conditions of AHT, DM, and DKD. The highest number of sentences was recorded for the discharge and patient-centered letters for DKD, while the fewest sentences were found in the patient-centered letters on DM.

Table 2: Quantitative Analysis of Letter Structure.

Data are presented as median (interquartile range) or absolute numbers.

	Overall	AHT	DM	DKD
Number of Sentences				
• Discharge letter	273	83	92	98
• Patient letters	860	278	271	311
Sentences per Letter				
• Discharge letter	92	83	92	98
• Patient letters	56 (54–59)	55 (52–57)	54 (53–58)	58 (56–65)
Words per Sentence				
• Discharge letter	8 (3–13)	9 (4–14)	8 (3–13)	7 (3–13)
• Patient letters	15 (8–20)	15 (9–21)	14 (9–21)	14 (8–19)

2. *Learning Objectives in GPT-4-Generated Patient-centered Letters*

A detailed illustration of the learning objectives addressed in the patient-centered letters is provided in **Figure 2**. Of 72 learning objectives (24 per medical condition across 3 conditions), 57 were identified and transitioned in the majority (≥3 of 5 letters per disease). However, no patient-centered letter included all the learning objectives present in the corresponding discharge letter. There was no significant difference in the coverage of learning objectives between the three different diseases (AHT, DM, and DKD).

Figure 3 illustrates the learning objectives categorized according to Bloom`s taxonomy and content field. Notably, the figure reveals that learning objectives classified under Bloom`s category Remember were more frequently depicted in the patient-centered letters compared to those under Understand.

Some learning objectives were only partially addressed in the majority (≥3 of 5 letters) of patient-centered letters (5 of 72). Examples of these partially addressed learning objectives are provided in **Table 3**. The missing information in these cases could be categorized into different areas (e.g., responsibility, frequency, dosage). However, no systematic categorical errors were identified in the analysis of the partially addressed learning objectives.

Of the 72 learning objectives, 11 were completely omitted in the majority (≥3 of 5 letters) of the patient-centered letters (**Supplementary Table 2**). Of these, only 2 belonged to the

Remember category, while the majority fell under the Understand category. Furthermore, most of the missing learning objectives (6 of 11) related to the field “prevention of complications”. In contrast, learning objectives related to “lifestyle” and “organizational” aspects were more frequently included in the patient-centered letters by GPT-4.

Table 3: Analysis of Incomplete Learning Objectives
Examples of learning objectives that have been partially listed in the patient-centered letters including recommendations for further prompt engineering.

Content field	Category	Examples	Recommendations for Prompt Engineering
Organizational	Lack of connection to the underlying disease	"Additionally, you should perform a special urine collection test at home, where your urine will be collected over a 24-hour period to search for additional hormones that may indicate other diseases."	(See idea about providing structure below)
	Information is missing that would enable specific actions:		
	• responsibility	"Every three months, one should have a blood test called HbA1c, which shows the average blood sugar level over the past few months."	Clearer definition of the context in which the patients operate (here, the German healthcare system). This makes the prompt longer, potentially hurting performance elsewhere.
	• frequency	"Eye doctor visits: Have your eyes checked regularly to ensure that your vision is not affected by diabetes."	
Medication	Information missing:		
	• dosage	"If your levels are above 150 mg/dl before breakfast on three consecutive days, you should take a bit more insulin."	Provide a clear expected structure for an action point: 1. required action 2. how to judge / measure success 3. what to do in case of bad measurements 4. why this is important 5. what happens if unchecked Provide one example of this expected structure
	• medication	"It has also been found that your cholesterol level is too high, and you have been started on a medication that can additionally reduce the risk of heart problems."	
Prevention of complications	Information missing:		
	• threshold	"If the pressure rises very high, you should go to the doctor immediately."	
	• recommended action for side effects	"You should watch for muscle pain, as this can be a side effect of the medication."	
Lifestyle/Disease management	Information is missing that would support specific actions		
	• target value	"You should also adjust your diet, eat less salt, and make sure you do not weigh too much."	
	• frequency	"Additionally, you have been advised to exercise regularly to improve your overall health."	

Supplementary Table 2: Qualitative Analysis of Missing Learning Objectives
This table illustrates learning objectives that were not included in the majority of patient-

centered letters (≥3 of 5 letters). Recommendations are provided on how the prompting can be refined to address the missing information

Bloom category	**Common disease**	**Learning objective (information added for better understanding)**	**No of patient letters in which learning objective is missing (x/5)**	**Content field**	**Recommendations for Prompt Engineering**
Remember	Type 2 diabetes mellitus	being vigilant of palpitations, trembling or changes in consciousness	4	**Prevention of complications**	
		arrange follow up with primary care physician every 3 months	3	Organizational	
Understand	Arterial hypertension	(target blood pressure below 130/80 mmHg) prevents cardiovascular damage	5	**Prevention of complications**	The idea about providing expected structure for an action point should also help here to some extent. This structure would need to include a subcategory that explains the “why”. An example of such a structure in the prompt which includes information from the Understand category may help further.
		(discontinue ramipril and amlodipine) included in the new fix-dose combination	4	Medication	
		(inform doctors about his/her allergy to codeine and Azithromycin) prevent accidental prescription of these substances	3	**Prevention of complications**	
		(ambulatory sleep apnea diagnostics sleep apnea) might be associated with hypertension	3	Organizational	
	Type 2 diabetes mellitus	(regular blood sugar checks) help guide insulin therapy	4	Disease Management/ Lifestyle changes	
		(being vigilant of palpitations, trembling or changes in consciousness) these are symptoms of hypoglycemia	4	**Prevention of complications**	
		(dosage of insulin based on blood sugar levels) prevents hyper- or hypoglycemic episodes	3	**Prevention of complications**	
	Diabetic nephropathy	(reducing Torasemide when a balanced volume status is achieved) knows that volume deficit can lead to kidney injury	5	**Prevention of complications**	
		(Reduce Metformin) needs to be dose-adjusted according to kidney function	4	**Prevention of complications**	

3. Indicators of Patient-Centricity

The Flesch Reading Ease test scores for both the discharge letters and the patient-centered letters are shown in **Figure 4**. The patient-centered letters scored around 60, which is equivalent to a “standard” reading level [26]. In contrast, the discharge letters scored around 40, which reflects a “difficult” reading level. As illustrated in Figure 4, the patient-centered letters have an LIX score of approximately 50, while discharge letters scored around 60, indicating greater difficulty in reading. The patient-centered letters also scored lower than the discharge letters on the RIX, demonstrating better readability.

Abbreviations were used in 18% (n=49) of the discharge letter sentences, compared to 5% (n=43) in the patient-centered letters. The abbreviation rate remained consistent across all patient letters (AHT/DM/DKD: n=14/14/15) at 5% for all cases.

Repetitions were found in 13% (n=37) of the discharge letter sentences and 24% (n=207) of the patient-centered letter sentences. The number of repetitions varied slightly across patient-centered letters (AHT/DM/DKD: n=60/75/72 (22%/28%/23%)).

Medical terms appeared in 60% (n=165) of the discharge letter sentences but in only 15% (n=132) of the patient-centered letters. Of the medical terms used in patient-centered letters, 92% (n=121) were explained.

Patients were directly addressed in 58 % (n= 502) of the sentences in the patient-centered letters. This varied slightly between the different patient-centered letters (AHT/DM/DKD n=151/149/202 (54%/55%/65%)).

4. Medical Correctness, Case-Specific Relevance and Source of Information

Of the patient letters, 756 sentences (88%) were medically correct, while 31 (4%) were medically incorrect. The rate of medically correct sentences was comparable among the patient letters for AHT (90%, n= 250), DM (89%, n=240) and DKD (86%, n=266). Medical errors could be found in 3% (n=21) of the sentences in the patient-centered letters on AHT, in 4% (n=21) in the letters on DM, and in 5% (n=31) in the letters on DKD (**Figure 6**).

Regarding case-specific relevance, 88% (756 of 860) of the sentences in the patient-centered letters were very relevant, while 8% (72 of 860 sentences) were rather relevant. In contrast, 4% (32 sentences) were not relevant. The proportion of highly relevant sentences was similar across the patient letters on AHT, DM, and DKD (**Figure 5**).

The information given in the patient-centered letters was derived from the discharge letters in 72% (616 of 860) sentences. Additional medical information was provided by GPT-4 in 24% (207 sentences). However, these proportions varied among the specific cases. The patient-centered letters on AHT were based on the discharge letters in 83% of the sentences, compared to 70% for DM and 63% for DKD. Further details are shown in **Figure 5.**

A qualitative analysis of the medical errors conducted following a thematic analysis based on Braun and Clarke can be found in **Supplementary Table 3**.

Supplementary Table 3: Qualitative Analysis of Medical Errors

Medical errors made by GPT-4 were categorized and analyzed based on content area. The errors were classified into different error categories: imprecise, incorrect, incomplete, presumptive.

Content field (counts)	Content	Unit of information, patient letter	Source (discharge letter)	Error category	Possible causes
Disease	Therapy	"Your blood sugar	"HbA1c checks	imprecise	Mixing of

management/ Lifestyle (6)	success monitoring/ Type of examination	should be checked every three months to ensure that the medications are working properly."	every three months."		different therapy monitoring examinations
	Therapy success monitoring/ Frequency of examination	"Monitor your blood sugar levels closely to ensure they are not too high or too low."	"During the therapy, it is recommended to perform fasting blood sugar checks once or twice a week to guide treatment management."	imprecise	
	Dietary measures	"Diet: Be sure to reduce your intake of salt, protein, and potassium to ease the strain on your kidneys."	"We recommend following a Mediterranean diet for this purpose."	incorrect (for this stage of the disease)	Recommendation not appropriate for this stage of the disease
Medical knowledge (8)	Explanation/ Definition of a disease	"The HbA1c value, which indicates the average blood sugar level over the past few weeks, is at 14.1%, which is also very high (normal is under 6%)."	No information available	incorrect	
	Explanation/ Definition of a disease	"Previously, there were issues with the esophagus just after the stomach due to acid (reflux esophagitis)."	"History of reflux esophagitis"	incorrect	
	Deduction of the most likely risk factor based on the disease	"The family also has a history of blood pressure issues, and his father had a stroke: This could indicate that Mr. Raser is at an increased risk for cardiovascular diseases."	"CVRF: Positive family history of a stroke in the father at the age of approximately 50-60 years."	presumptive	Independent inference of the most likely risk factor
	Interpretation of a test result	"They also tested to see if your pancreas has any issues, which it does not."	"An ultrasound of the pancreas showed no evidence of a mass."	presumptive	Improper generalization (due to intended simplification?)
medication (7)	Route of administratio	"Insulin glargine: Take 14 'pumps' of	"Insulin glargine 100 U/mL: 14	incorrect	Incomplete medication plan

	n	this special diabetes medication every morning to keep your blood sugar under control."	units at 8 a.m., NEW."		(missing method of administration) in the letter and improper simplification (possibly due to intended simplification?)
	Time of administratio n	"Medications: Ramipril/Amlodipine: Now take these blood pressure tablets twice daily."	"Ramipril/ Amlodipine 5 mg/ 5 mg: 2 – 0 – 0, INCREASED, previously 1 – 0 – 0"	incorrect	
	Medication adjustment	"You take this tablet every morning, and it replaces your previous medications, namely Ramipril and Amlodipine at a lower dosage."	"Ramipril/ Amlodipine/HCT 10/10/25 mg: 1 – 0 – 0, NEW Ramipril 5 mg: 1 – 0 – 1, DISCONTINUE D Amlodipine 10 mg: 1 – 0 – 0, DISCONTINUE D" "...thus, we recommend a therapy adjustment as indicated above..."	incomplete	
Prevention of complications (3)	Sick day rules	"If you develop a fever or an infection, you should temporarily discontinue a specific diabetes medication to avoid severe side effects.”	"In the event of an infectious disease with fever, the therapy with the SGLT2 inhibitor, as well as the therapy with metformin, should be paused due to the increased risk of ketoacidosis or lactic acidosis."	incomplete	Multidimensional problem which requires adherence to a strict process chain.
	Allergy	"It is important that you see a doctor immediately at the first sign of an infection, especially since you have an allergy to penicillin."	"Please exercise increased vigilance for penicillin allergy during a doctor's visit in case of infection."	incorrect	

Organizational (3)	Incorrect assignment of time	"Mr. Süss was treated as an inpatient, staying overnight in a hospital on October 8, 2023, after visiting a general practitioner."	"Admission: 04/10/2023 / Discharge: 08/10/2023"	incorrect	

No discernible pattern was observed in the errors within patient-centered letters. Some errors resulted from imprecision or incomplete information, while others were due to incorrect assumptions made by GPT-4. In general, most of the sentences (89%, n=182) added by GPT-4 were medically correct. However, of the medically incorrect sentences, the majority (71%, n=22) contained information provided solely by GPT-4.

Discussion

This study demonstrates the potential of GPT-4s to transform discharge letters into more readable, patient-centered letters. Even if this is in line with previous studies, showing GPT-4's ability to generate patient-centered letters based on discharge letters or shorthand clinical instructions [16, 27], our study went beyond the proof of concept for patient letter generation and focused on communicating relevant information from the patient-safety perspective. To our knowledge, this is the first study to analyze this view in AI-generated patient-centered letters.

Automated Learning Objective Identification and Transformation of Treatment and Patient Safety-Relevant Information

Our main focus was on the identification and transfer of information relevant to patient safety, defined as "learning objectives," in GPT-4-generated patient-centered letters. Overall, GPT-4 effectively identified and transformed 56 of 72 learning objectives, with 5 partially addressed and 11 missed entirely.

A key finding is that no significant differences were observed between the three diseases (AH, DM and DKD) with regard to the quality of GPT-4-based learning-objective identification and transformation.

Another very important finding is that GPT-4 fully identified and transformed 77.8% of the learning objectives from the discharge letters. In 6.9% of the cases, GPT-4 identified the learning objectives of the discharge letters that are relevant for treatment and patient safety but only transformed them into the patient-centered letters with the omission of individual aspects. In 15.3% of the cases, GPT-4 did not identify any learning objectives and therefore did not transform them and add them to the patient-centered letters.

In line with previous studies in the medical field, we found that the phenomenon of omitting relevant key information in the medical context, as described in AI research, continues to exist[16, 27].

Since the phenomenon of the partial omission of relevant key information in the medical context is highly relevant to the quality and safety of patient care, we have examined this phenomenon in detail in the medical context, for the first time to our knowledge, in the present study.

From our point of view, it is particularly interesting that we were able to show that omissions of GPT-4 occurred more frequently with more complex requirements in the medical context. These were mainly learning objectives that require a deeper understanding (Bloom category: Understand), compared to simpler learning objectives (Bloom category: Remember). Basic topics such as 'lifestyle' and 'organizational' aspects of self-management of one's own health (example: regularly attending follow-up examinations every 3 months) were more frequently recognized by GPT-4 and transformed into patient-friendly language. More complex content such as 'prevention of complications' (example: interactions between certain foods and medications), on the other hand, was addressed less frequently by GPT-4. In summary, this phenomenon was particularly evident in the processing and transformation of complex logical structures with multiple dependencies.

To address this challenge in a needs-based way, "few-shot learning," in which the AI model is guided by several concise examples, and the use of "chain-of-thought" prompts, which break down complex, multilevel problems into a series of intermediate steps and thus improve the AI's performance at high complexity, could be potentially promising strategies to optimize GPT-4's translation of complex medical information into more easily understandable statements of fact and directions to follow [28, 29].

In our case, the "chain-of-thought" prompting would need to include an initial step that clarifies to the AI why the patient should adhere to the respective learning objective. An example of such a structure could be to first identify the content relevant to the patient (e.g., 'pay attention to interactions'). The next step could be to ask the AI to break this content down into individual pieces of information (e.g., medication, diet, kidney function). In the subsequent processing step, the AI should put these pieces of information into the correct order. Only in the subsequent step should this order be checked for existing dependencies between the individual information units and adjusted if necessary. This could help to prevent the omission of complex learning objectives that are important for patient safety and quality of care. It should be noted that this approach extends the AI-prompting and may affect the performance of the AI elsewhere. Therefore, this potentially promising strategy should be specifically addressed in future studies to improve the complete translation of complex medical information by GPT-4 into more easily understandable information for the patients.

Regarding the learning objectives that were only partially transformed by GPT-4 from the discharge letters to the patient letters, we could not identify a clear connection with the content. In previous studies, medical notes were used to demonstrate GPT-4's ability to extract specific information with a high degree of accuracy. However, in these studies, GPT-4 focused exclusively on the task of information extraction.

In contrast, our study required a multitasking prompt consisting of the identification of learning objectives and the subsequent transformation of these medical facts in the correct context and in simple language for the creation of patient-centered letters. In analogy to the findings regarding the complete omission of complex learning objectives, the challenge for AI in processing and reproducing complex logical structures with multiple dependencies was also evident here.

Particularly challenging for GPT-4 seemed to be learning objectives from which no immediate and simple logical pattern can be derived. This could possibly be due to multitasking prompts (identification, summarization, and simplification of medical information) and the overall high complexity of medical issues when transformed into a barrier-free language that patients can understand. Another approach for the prompt could therefore be to implement 'action points' to clearly structure the learning objectives into required actions, success criteria, and possible consequences. This strategy should be further investigated in future studies.

Quality of Medical Information after Transformation by GPT-4

In the present study, it was also particularly important to us to examine the medical accuracy of the medical information transformed by GPT-4 in the patient-centered letters, since this is crucial for patient safety and quality of care and any inaccuracies must therefore be assessed as a risk for the use of LLMs in a medical context.

Our results show that the patient-centered letters generated by GPT-4 exhibited a high degree of medical correctness. The low error rate observed in our study could be due to the detailed medical information base of the discharge letters that GPT-4 was able to access during the transformation process. This extensive dataset enabled the GPT-4 model to transform the medical information with a high medical accuracy into patient-centered language. Previous studies have shown that, in addition to the information omissions, the type of task (e.g., text summary) and the specific field of application (e.g., medical writing, question-answer format) influence the accuracy of LLM transformation[30]. It is noteworthy that our study was able to demonstrate a substantial improvement in medical correctness and thus in the quality of the results compared to previous work that examined different scenarios for the use of LLMs in the medical context[27, 31].

A problem that should not be neglected when transforming medical information by LLMs is the phenomenon of “hallucinations,” in which coherent and grammatically correct, but factually wrong or misleading information is generated [32, 33]. These hallucinations can endanger patient safety and the quality of care. In previous studies, a hallucination rate of up to 40% was found, whereas we only encountered this phenomenon in isolated cases in our study [34]. Nevertheless, despite the downward trend, the risk remains significant with the continuous development of LLMs, as the following example shows:

> Discharge letter: “(...) An ultrasound of the pancreas showed no signs of a mass. (...)”

> Patient-centered letter: “(…) they also tested to see if your pancreas has any issues, which it does not (…).”

Here, GPT-4 makes a general statement about the unremarkable ultrasound of the pancreas, although medically only a malignant disease was excluded and other potential pathologies were not examined at all. Although these types of errors, caused by the hallucination phenomenon, occurred only very rarely, they nevertheless would pose a significant risk to patient safety and quality of care. Therefore, despite its low frequency, this phenomenon must still be considered a significant limitation to the use of LLMs in a medical context and must be specifically addressed in future studies.

Patient-Centricity

In terms of readability and patient-centricity, measured by standard readability scores, word and sentence counts, and the use of specialized terminology, explanations, abbreviations, and direct patient address, the GPT-4-generated patient-centered letters achieved a remarkably higher standard compared to the conventional discharge letters.

This is consistent with previous studies that analyzed AI-generated patient letters, albeit with less comprehensive methodologies, and showed a high readability ranging from sixth [16] to ninth [27] grade on the school reading level. It should be emphasized that these studies generally used the Flesch Reading Ease, the Flesch-Kincaid Reading Level or the PEMAT score for the readability assessments [16, 35].

The Flesch Reading Ease is one of the most commonly used tools for evaluating readability in medical literature and ranges from 0 (unreadable) to 100 (very easy to read)[22]. However, a major limitation of this approach is that the Flesch Reading Ease only refers to sentence and word length and thus does not adequately consider the necessary level of medical knowledge by the patients or the intensity of the use of medical terms.

From the authors' point of view, these factors, in conjunction with addressing patients directly, are crucial for the perception and understanding of the discharge letter by patients. In particular, previous studies have shown that avoiding abbreviations and explaining medical terms in layman's terms significantly improved the comprehensibility of discharge letters from the patient's point of view [36] and that medical letters are demonstrably better received by patients if they are addressed personally in the letter [37].

For the present study, we therefore implemented additional methodological approaches to analyze the patient-centered transformation of medical information by GPT-4 for the first time. These advanced methods not only considered the use of medical terms and abbreviations but also their explanations and the occurrence of repetitions.

It is noteworthy that the patient-centered letters, transformed by GPT-4, showed a significantly higher level of patient-centricity than conventional discharge letters, even when these advanced parameters, which were examined for the first time, were included. The GPT-4-generated patient-centered letters contained significantly fewer medical terms and abbreviations, the medical terms were explained in simple language, and the patients were addressed directly in the letter.

Limitations

Our study has several limitations. First, the fact that GPT-4's generation of patient-centered letters was based solely on the extensive prompting technique. In the long term—especially with the increasing availability of further LLMs [38, 39]—a fine-tuning (both in general and domain-specific terms) and the potential application of specialized LLM models should be incorporated into the process of creating patient-centered letters. Secondly, we focused our investigation on major widespread diseases that occur frequently but do not include all medical specialties. As a result, our findings may not be fully transferable to all medical specialties. Another limitation is the exclusive application of the LLM “GPT-4”, which does not allow us to make statements about the performance of other LLMs with regard to the creation of patient-centered letters based on discharge letters.

Furthermore, we used a new multidimensional approach to assess patient-centricity of the GPT-4-based communication of medical information. This approach considered various factors such as the use of medical terms, abbreviations, and repetitions. Although these variables are objectively measurable and part of common readability standards, they have not been sufficiently validated in the context of LLM-based communication. Therefore, further studies are needed that explicitly examine how patients perceive and understand the patient letters created by LLMs, and how this innovative form of patient-centered communication affects patient empowerment and self-management of their own health. Furthermore, the perspective of the treating physicians on this new, in our view promising, patient-centered communication was not examined. This aspect should also be considered in future studies.

Conclusion

In summary, our study shows that GPT-4 has the potential to make discharge letters significantly more patient-centered. Although we used a detailed prompting technique in our study and GPT-4 overall shows a high degree of medical correctness when transforming discharge letters into more patient-centered letters, it is not yet fully suitable for use in patient care without content review by medical professionals, particularly due to the described omissions and hallucinations. Despite the already good readability and patient orientation, even an advanced language model like GPT-4 did not fully consider all information relevant to patient safety and quality of care in the patient-centered letters. Further improvements in the prompting technology and targeted development of the language models for the medical field could help to minimize these limitations. If these challenges are overcome, GPT-4 could have increased potential to support doctors in patient-centered communication and to improve patients' understanding of their medical condition. This would be a significant step toward greater patient safety and quality of care.

Statements and Declarations

Conflict of interests

The authors declare that they have no conflict of interest.

Ethics approval

This study was approved by the local Ethics Committee (778/2023BO2) of Tübingen University hospital.

Informed consent to participate

Not required.

Informed consent for publication

Not applicable.

Funding

This study was supported by the Open Access Publishing Fund of the University of Tübingen.

Acknowledgments

We wish to thank the Open Access Publishing Fund of the University of Tübingen for supporting our study.

References

1. Osterberg L, Blaschke T. Adherence to medication. N Engl J Med. 2005 Aug 4;353(5):487-97. PMID: 16079372. doi: 10.1056/NEJMra050100.
2. Sabaté E. Adherence to long-term therapies: evidence for action: World Health Organization; 2003. ISBN: 9241545992.
3. DiMatteo MR, Haskard-Zolnierek KB, Martin LR. Improving patient adherence: a three-factor model to guide practice. Health Psychology Review. 2012;6(1):74-91. doi: 10.1080/17437199.2010.537592.
4. Longtin Y, Sax H, Leape LL, Sheridan SE, Donaldson L, Pittet D. Patient participation: current knowledge and applicability to patient safety. Mayo Clin Proc. 2010 Jan;85(1):53-62. PMID: 20042562. doi: 10.4065/mcp.2009.0248.
5. Schwarz CM, Hoffmann M, Schwarz P, Kamolz LP, Brunner G, Sendlhofer G. A systematic literature review and narrative synthesis on the risks of medical discharge letters for patients' safety. BMC Health Serv Res. 2019 Mar 12;19(1):158. PMID: 30866908. doi: 10.1186/s12913-019-3989-1.
6. Roy CL, Poon EG, Karson AS, Ladak-Merchant Z, Johnson RE, Maviglia SM, et al. Patient safety concerns arising from test results that return after hospital discharge. Ann Intern Med. 2005 Jul 19;143(2):121-8. PMID: 16027454. doi: 10.7326/0003-4819-143-2-200507190-00011.
7. Forster AJ, Murff HJ, Peterson JF, Gandhi TK, Bates DW. The incidence and severity of adverse events affecting patients after discharge from the hospital. Ann Intern Med. 2003 Feb 4;138(3):161-7. PMID: 12558354. doi: 10.7326/0003-4819-138-3-200302040-00007.
8. Harris E, Rob P, Underwood J, Knapp P, Astin F. Should patients still be copied into their letters? A rapid review. Patient Educ Couns. 2018 Dec;101(12):2065-82. PMID: 30420045. doi: 10.1016/j.pec.2018.06.014.
9. Smolle C, Schwarz CM, Hoffmann M, Kamolz LP, Sendlhofer G, Brunner G. Design and preliminary evaluation of a newly designed patient-friendly discharge letter - a randomized, controlled participant-blind trial. BMC Health Serv Res. 2021 May 12;21(1):450. PMID: 33975590. doi: 10.1186/s12913-021-06468-3.
10. Nguyen DL, Ambinder EB, Jones MK, Hill G, Harvey SC. Improving Patient Comprehension of Screening Mammography Recall Lay Letters. J Am Coll Radiol. 2019 Dec;16(12):1669-76. PMID: 31199890. doi: 10.1016/j.jacr.2019.05.029.
11. Wolff J, Auber G, Schober T, Schwär F, Hoffmann K, Metzger M, et al. Work-Time Distribution of Physicians at a German University Hospital. Dtsch Arztebl Int. 2017 Oct 20;114(42):705-11. PMID: 29122102. doi: 10.3238/arztebl.2017.0705.
12. Garabet R, Mackey BP, Cross J, Weingarten M. ChatGPT-4 Performance on USMLE Step 1 Style Questions and Its Implications for Medical Education: A Comparative Study Across Systems and Disciplines. Med Sci Educ. 2024 Feb;34(1):145-52. PMID: 38510401. doi: 10.1007/s40670-023-01956-z.
13. Cascella M, Montomoli J, Bellini V, Bignami E. Evaluating the Feasibility of ChatGPT in Healthcare: An Analysis of Multiple Clinical and Research Scenarios. J Med Syst. 2023 Mar 4;47(1):33. PMID: 36869927. doi: 10.1007/s10916-023-01925-4.
14. Johnson-Laird P. Deductive reasoning. Wiley Interdiscip Rev Cogn Sci. 2010 Jan;1(1):8-17. PMID: 26272833. doi: 10.1002/wcs.20.
15. Eppler MB, Ganjavi C, Knudsen JE, Davis RJ, Ayo-Ajibola O, Desai A, et al. Bridging the Gap Between Urological Research and Patient Understanding: The Role of Large Language Models in Automated Generation of Layperson's Summaries. Urol Pract. 2023 Sep;10(5):436-43. PMID: 37410015. doi: 10.1097/UPJ.0000000000000428.
16. Zaretsky J, Kim JM, Baskharoun S, Zhao Y, Austrian J, Aphinyanaphongs Y, et al. Generative Artificial Intelligence to Transform Inpatient Discharge Summaries to Patient-Friendly Language and Format. JAMA Netw Open. 2024 Mar 4;7(3):e240357. PMID:

38466307. doi: 10.1001/jamanetworkopen.2024.0357.
17. Mac O, Ayre J, Bell K, McCaffery K, Muscat DM. Comparison of Readability Scores for Written Health Information Across Formulas Using Automated vs Manual Measures. JAMA Netw Open. 2022 Dec 1;5(12):e2246051. PMID: 36508219. doi: 10.1001/jamanetworkopen.2022.46051.
18. Herrmann-Werner A, Festl-Wietek T, Holderried F, Herschbach L, Griewatz J, Masters K, et al. Assessing ChatGPT's Mastery of Bloom's Taxonomy Using Psychosomatic Medicine Exam Questions: Mixed-Methods Study. J Med Internet Res. 2024 Jan 23;26:e52113. PMID: 38261378. doi: 10.2196/52113.
19. Bloom B FE, Engelhart M, Hill W, Krathwohl D. Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain. London: Longman Green & Co.; 1956.
20. Krathwohl DR. A Revision of Bloom's Taxonomy: An Overview. Theory Into Practice. 2002;41(4):212-8.
21. Rooney MK, Santiago G, Perni S, Horowitz DP, McCall AR, Einstein AJ, et al. Readability of Patient Education Materials From High-Impact Medical Journals: A 20-Year Analysis. J Patient Exp. 2021;8:2374373521998847. PMID: 34179407. doi: 10.1177/2374373521998847.
22. Jindal P, MacDermid JC. Assessing reading levels of health information: uses and limitations of flesch formula. Educ Health (Abingdon). 2017 Jan-Apr;30(1):84-8. PMID: 28707643. doi: 10.4103/1357-6283.210517.
23. Björnsson CH. Läsbarhet. Stockholm: Liber; 1968.
24. Anderson RC, Freebody P. Reading comprehension and the assessment and acquisition of word knowledge. Advances in Reading/Language Research. 1983;2:231-56.
25. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology. 2006 2006/01/01;3(2):77-101. doi: 10.1191/1478088706qp063oa.
26. Spadaro DC, Robinson LA, Smith LT. Assessing readability of patient information materials. Am J Hosp Pharm. 1980 Feb;37(2):215-21. PMID: 7361793.
27. Ali SR, Dobbs TD, Hutchings HA, Whitaker IS. Using ChatGPT to write patient clinic letters. Lancet Digit Health. 2023 Apr;5(4):e179-e81. PMID: 36894409. doi: 10.1016/S2589-7500(23)00048-1.
28. Paranjape B, Lundberg SM, Singh S, Hajishirzi H, Zettlemoyer L, Ribeiro MT. ART: Automatic multi-step reasoning and tool-use for large language models. ArXiv. 2023;abs/2303.09014.
29. Wei J, Wang X, Schuurmans D, Bosma M, Ichter B, Xia F, et al. Chain-of-thought prompting elicits reasoning in large language models. Proceedings of the 36th International Conference on Neural Information Processing Systems; New Orleans, LA, USA: Curran Associates Inc.; 2024. p. Article 1800.
30. Dave T, Athaluri SA, Singh S. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. Front Artif Intell. 2023;6:1169595. PMID: 37215063. doi: 10.3389/frai.2023.1169595.
31. Eriksen AV, Möller S, Ryg J. Use of GPT-4 to Diagnose Complex Clinical Cases. Nejm Ai. 2023;1(1). doi: 10.1056/AIp2300031.
32. Preiksaitis C, Rose C. Opportunities, Challenges, and Future Directions of Generative Artificial Intelligence in Medical Education: Scoping Review. JMIR Med Educ. 2023 Oct 20;9:e48785. PMID: 37862079. doi: 10.2196/48785.
33. Masters K. Medical Teacher's first ChatGPT's referencing hallucinations: Lessons for editors, reviewers, and teachers. Med Teach. 2023 Jul;45(7):673-5. PMID: 37183932. doi: 10.1080/0142159X.2023.2208731.
34. Chen TC, Kaminski E, Koduri L, Singer A, Singer J, Couldwell M, et al. Chat GPT as a Neuro-Score Calculator: Analysis of a Large Language Model’s Performance on Various

Neurological Exam Grading Scales. World Neurosurgery. 2023 2023/11/01/;179:e342-e7. doi: https://doi.org/10.1016/j.wneu.2023.08.088.
35. Kirchner GJ, Kim RY, Weddle JB, Bible JE. Can Artificial Intelligence Improve the Readability of Patient Education Materials? Clin Orthop Relat Res. 2023 Nov 1;481(11):2260-7. PMID: 37116006. doi: 10.1097/CORR.0000000000002668.
36. Weetman K, Dale J, Scott E, Schnurr S. Adult patient perspectives on receiving hospital discharge letters: a corpus analysis of patient interviews. BMC Health Serv Res. 2020 Jun 15;20(1):537. PMID: 32539716. doi: 10.1186/s12913-020-05250-1.
37. Mahadavan L, Bird NJ, Chadwick M, Daniels IR. Prospective assessment of patient directed outpatient communication from a patient and general practitioner perspective. Postgraduate Medical Journal. 2009;85(1006):395-8. doi: 10.1136/pgmj.2008.068601.
38. Peng C, Yang X, Chen A, Smith KE, PourNejatian N, Costa AB, et al. A study of generative large language model for medical research and healthcare. npj Digital Medicine. 2023 2023/11/16;6(1):210. doi: 10.1038/s41746-023-00958-w.
39. Jiang LY, Liu XC, Nejatian NP, Nasir-Moin M, Wang D, Abidin A, et al. Health system-scale language models are all-purpose prediction engines. Nature. 2023 Jul;619(7969):357-62. PMID: 37286606. doi: 10.1038/s41586-023-06160-y.

Supplement:
Supplemental Material 1:

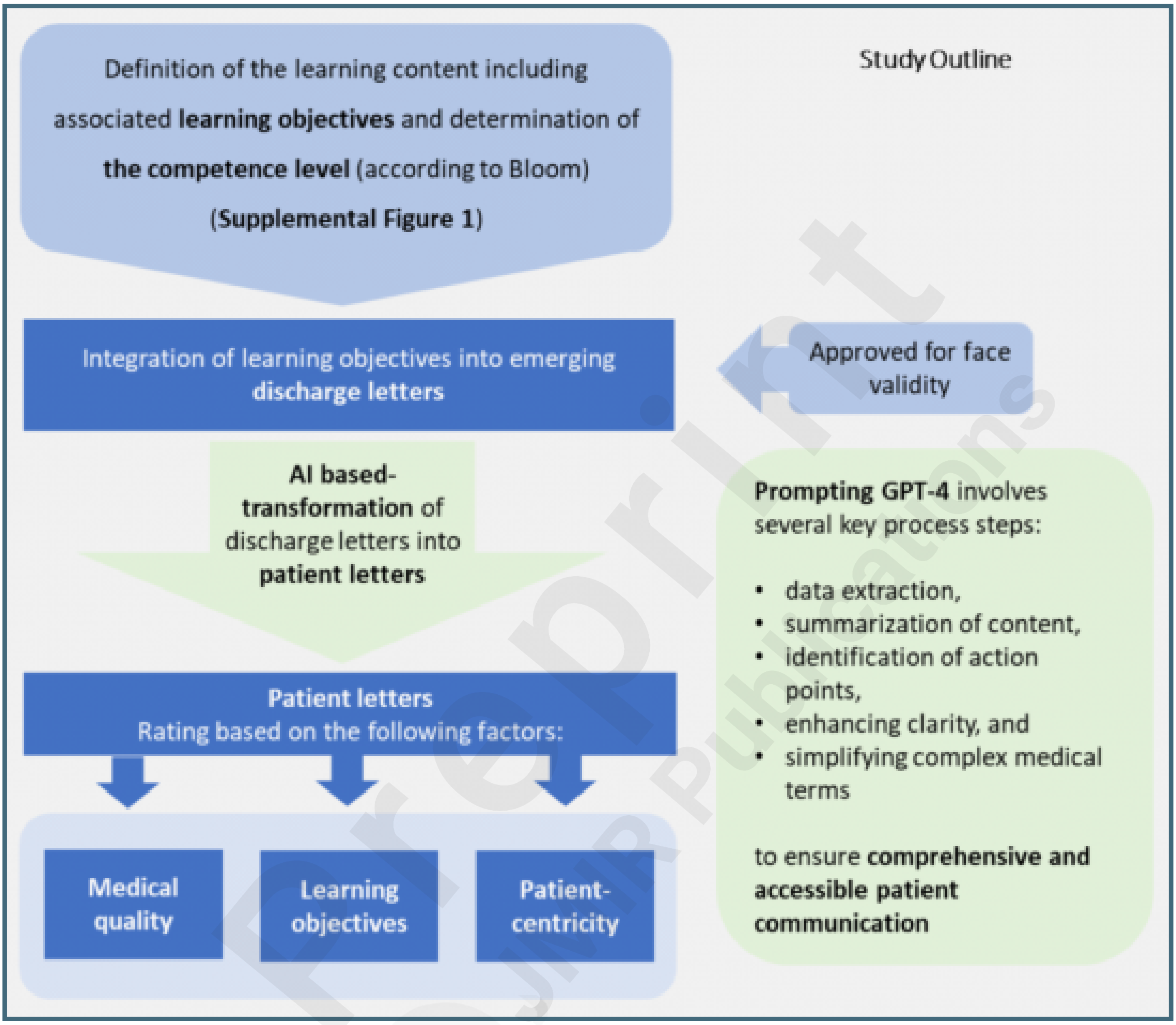
- Discharge letters
- Patient letters (examples)
- Table with learning objectives

Supplementary Figure 1: Development of Discharge Letters
Supplementary Figure 2: Prompt Development

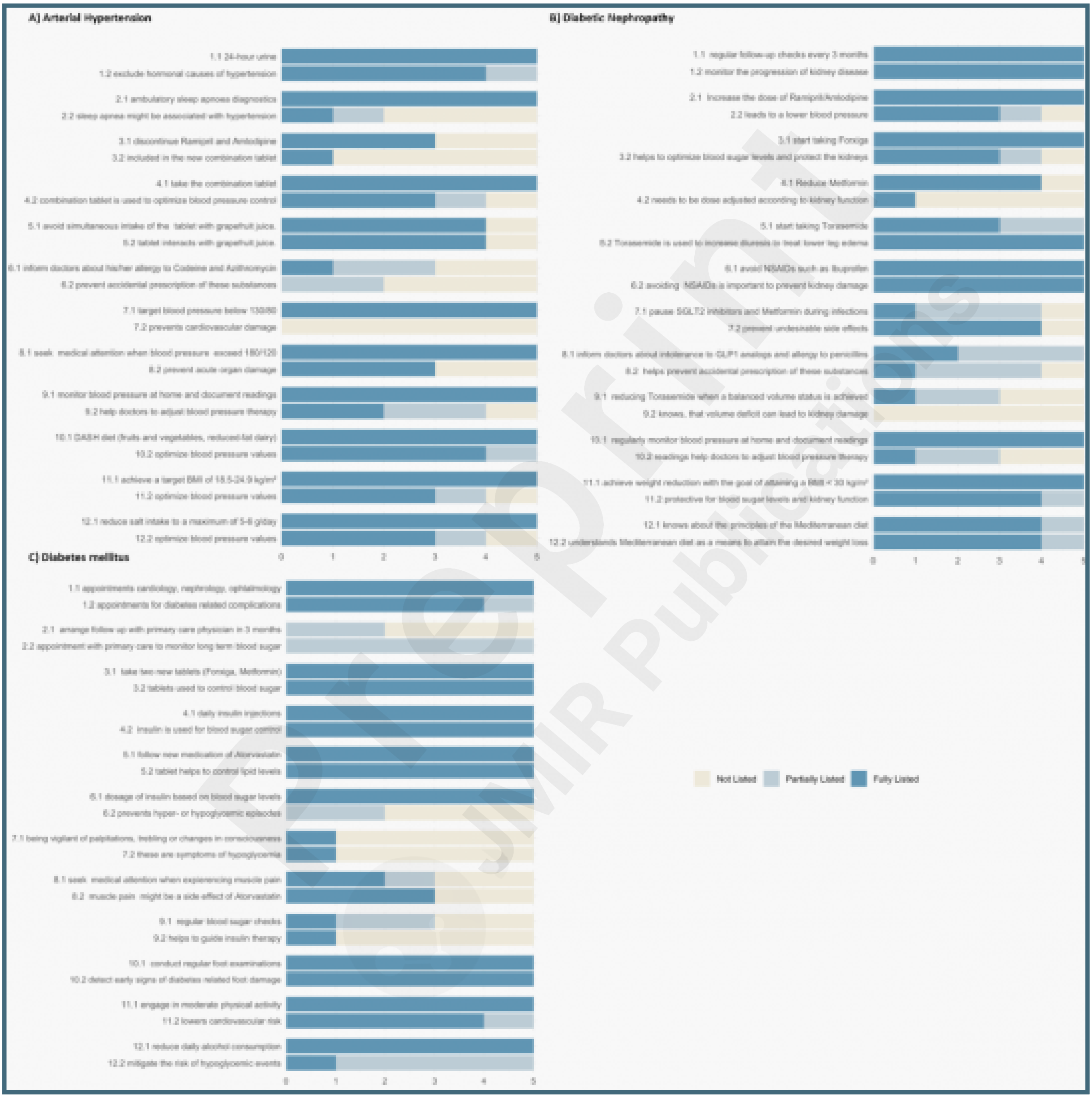
Supplementary Files

Figures

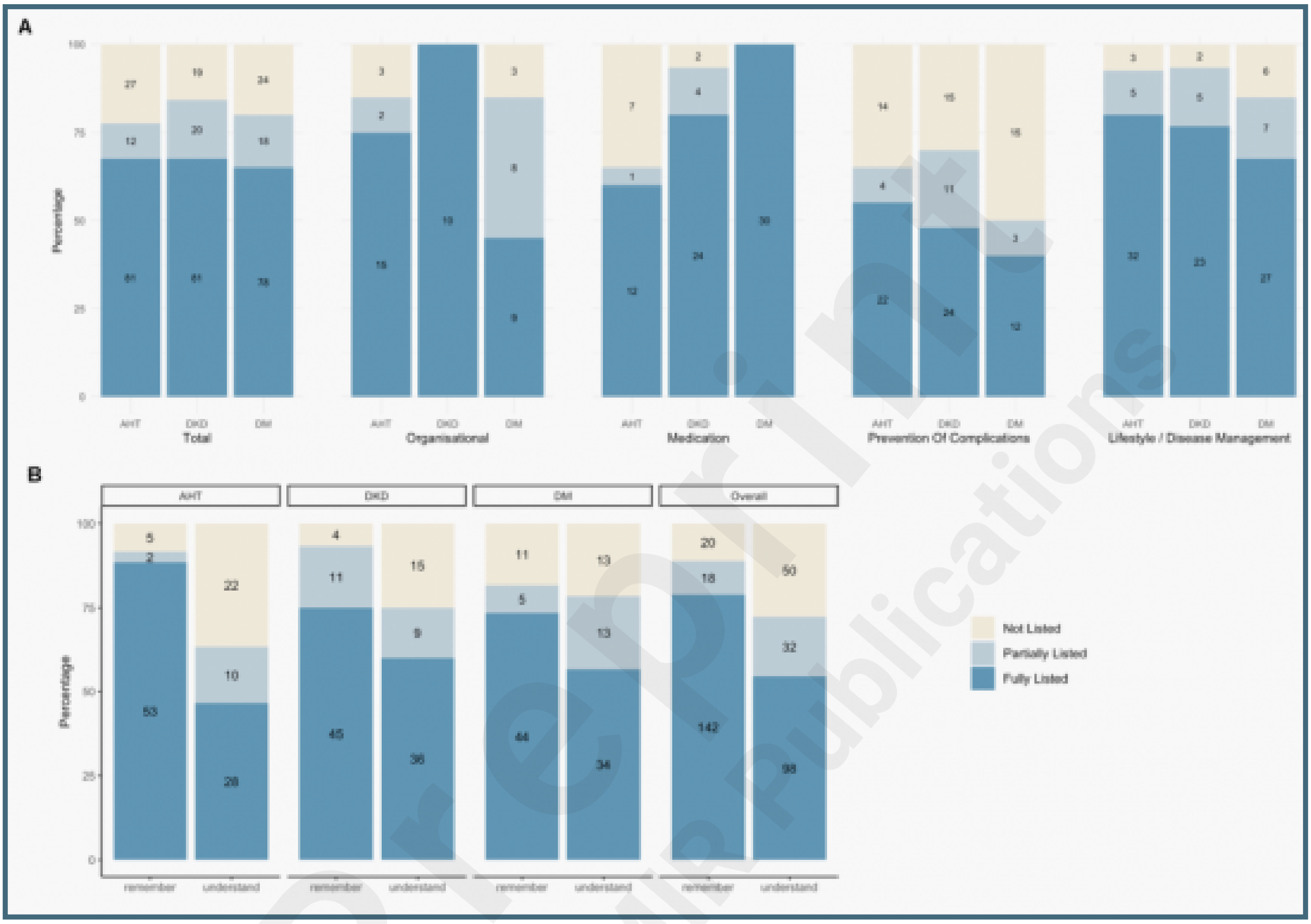
Comprehensive overview of the study’s structure and progression. Prompting steps are illustrated in the green box.


Study Outline
Definition of the learning content including associated **learning objectives** and determination of **the competence level** (according to Bloom) (**Supplemental Figure 1**)
Integration of learning objectives into emerging discharge letters
Approved for face validity
AI based-transformation of discharge letters into **patient letters**
Patient letters
Rating based on the following factors:
Medical quality
Learning objectives
Patient-centricity
Prompting GPT-4 involves several key process steps:
• data extraction,
• summarization of content,
• identification of action points,
• enhancing clarity, and
• simplifying complex medical terms
to ensure **comprehensive and accessible patient communication**

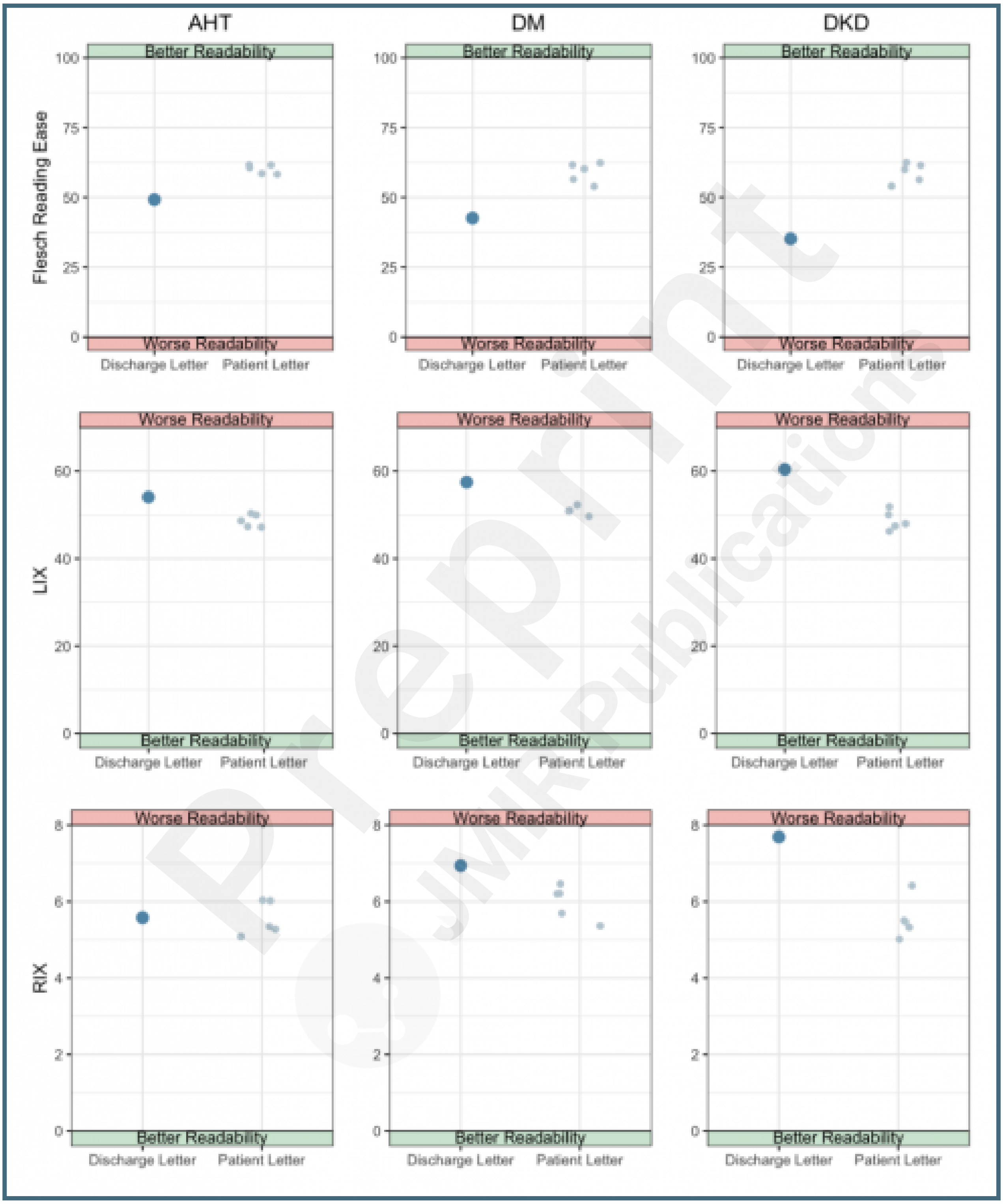
A–C) Illustration of the learning objectives described in the discharge letters and their representation in the patient-centered letters. The colors indicate the extent to which each learning objective has been addressed in the patient-centered letters (dark blue: fully listed, light blue: partially listed, yellow: not listed). The learning objectives are described on the Y-axis and the number of patient letters per disease on the X-axis.



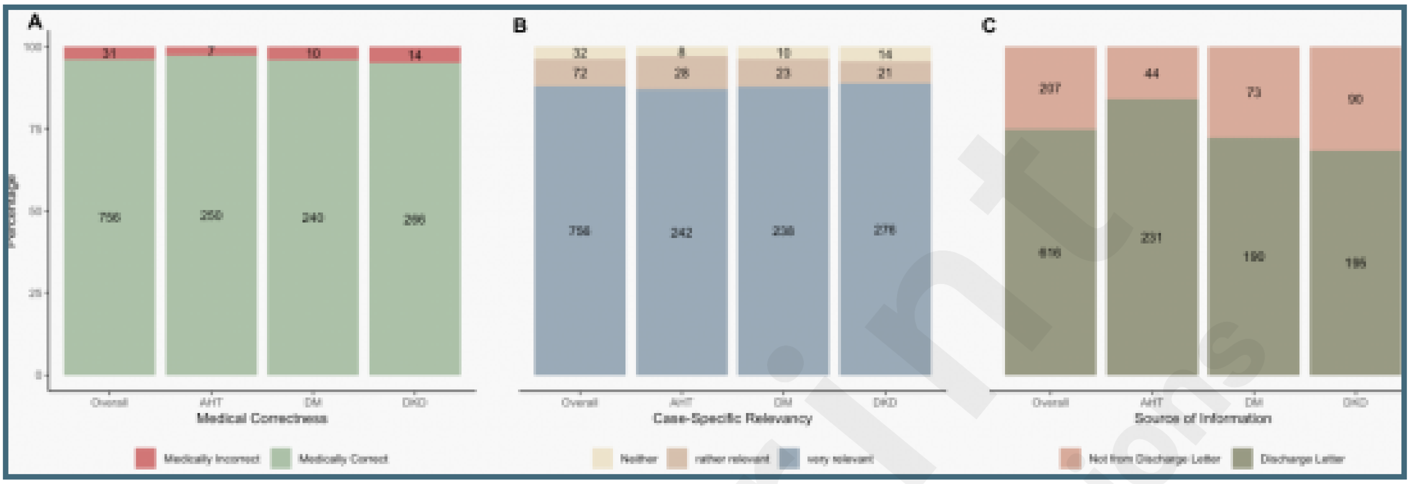
A) Classification of learning objectives according to the relevant content field (Organizational, Medication, Prevention of Complications, Lifestyle). B) Classification of learning objectives based on Bloom’s revised taxonomy (Remember, Understand). The colors indicate the extent to which the learning outcome is included in the patient letter (dark blue: fully included, light blue: partially included, yellow: not included).



The graph shows the readability scores for the discharge letters and the patient-centered letters. In the Flesch Reading Ease test, a higher score indicates a better readability (green, good readability), while in the LIX and RIX, a higher score represents lower readability (red, worse readability).



A) Illustration of medically correct (green) and incorrect (red) sentences in the patient-centered letters with the absolute number of sentences displayed within the bars. B) Representation of the case-specific relevance of the sentences in the patient-centered letters (blue: very relevant, brown: rather relevant, yellow: neither/nor relevant). The absolute number of sentences is shown in the bars. C) Proportion of sentences derived from the discharge letter (gray) and from GPT-4 (red) with the absolute number of sentences displayed within the bars.



Multimedia Appendixes

Supplementary Material: Discharge Letters, Patient Letters (examples) and Learning Objectives.
URL: http://asset.jmir.pub/assets/720de1fb5d64e355f67a664bbe87f86a.docx

Supplemental Figure 1: Development of Discharge Letters Illustration of the multistep process used for the development of discharge letters. Three common diseases were selected as the basis for the discharge letters. To assess GPT-4`s ability to identify critical medical information, 24 learning objectives were developed in alignment with Bloom's taxonomy. Examples are given in the figure.
URL: http://asset.jmir.pub/assets/57f35efea6e760eda1c8fb820c69cb1f.png

Supplemental Figure 2: Prompt Development Overview of the multistep process of GPT-4 prompting. The three tasks involved in the generation of the patient-centered letters are illustrated in the rectangles. The direction of the arrows indicates the sequential order of the individual steps.
URL: http://asset.jmir.pub/assets/d7975bf9135eaad98d2c6182eeef617a.png