WORK IN PROGRESS

Reframing Automation Benefits: A Human-Centered Approach to Expanding the Value of AI-Generated Reports in Healthcare

HUBERT DARIUSZ ZAJAÇ, University of Copenhagen, Copenhagen, Hovedstaden, Denmark

NIELS KVORNING TERNOV, Gentofte Hospital, Hellerup, Hovedstaden, Denmark

Open Access Support provided by:

Gentofte Hospital

University of Copenhagen

Published: 26 April 2025

Citation in BibTeX format

CHI EA '25: Extended Abstracts of the CHI Conference on Human Factors in Computing Systems
April 26 - May 1, 2025
Yokohama, Japan

Conference Sponsors:
SIGCHI

CHI EA '25: Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (April 2025)
https://doi.org/10.1145/3706599.3719979
ISBN: 9798400713958

Reframing Automation Benefits: A Human-Centered Approach to Expanding the Value of AI-Generated Reports in Healthcare

Hubert Dariusz Zając
Department of Computer Science
University of Copenhagen
Copenhagen, Denmark
hdz@di.ku.dk

Niels Kvorning Ternov
Department of Plastic Surgery
Herlev and Gentofte University Hospital
Herlev, Denmark
niels.kvorning.ternov@regionh.dk

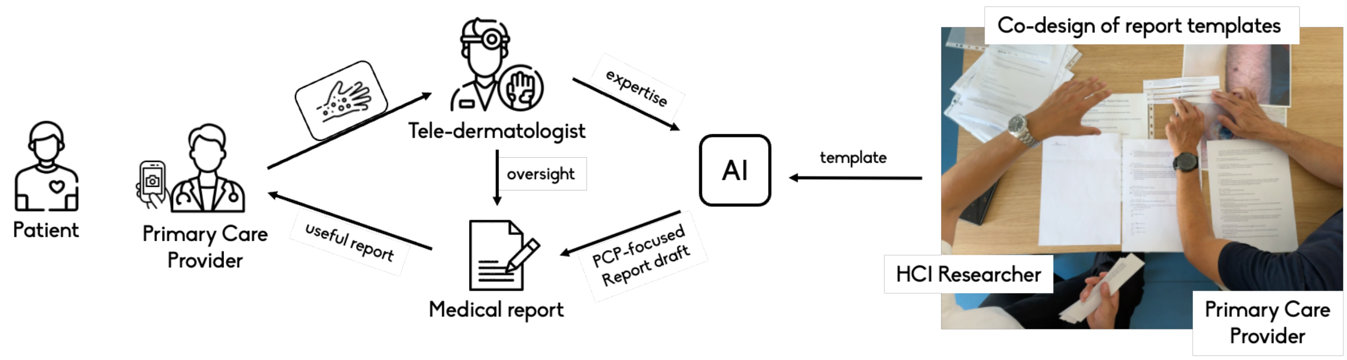

Figure 1: A co-design session with a general practitioner to establish a dermatology report template that reflects the preferences and informational needs of the recipients, providing a foundation for AI-generated reports.

Abstract

The use of Large Language Models (LLMs) for generating medical reports is being increasingly explored across medical specialties. However, these solutions often prioritize the perspectives of report authors, leaving the recipients and their needs outside of the scope of consideration. In this dermatology-focused study, we conducted co-design sessions with general practitioners (GPs) (N=12) in Denmark to establish their content preferences in dermatology reports and investigate unmet information needs in specialist-authored reports. We discuss using their preferences as input to generative AI to enhance the usefulness of medical reports. Such AI could support dermatologists by drafting reports and acting as a proxy for GP information needs, thus improving GP efficacy, enhancing patient outcomes, and reducing the overall burden on healthcare systems. Building on this study, we plan to: refine report structures with dermatologists and patients focusing on collaborative decision-making; and investigate non-chat genAI interaction space for automated reporting in safety-critical environments.

CCS Concepts

• **Human-centered computing → Collaborative and social computing theory, concepts and paradigms**; • **Computing methodologies → Artificial intelligence**.

ACM ISBN 979-8-4007-1395-8/25/04
https://doi.org/10.1145/3706599.3719979

Keywords

LLM, GenAI, AI, automated reporting, clinical reporting

ACM Reference Format:
Hubert Dariusz Zając and Niels Kvorning Ternov. 2025. Reframing Automation Benefits: A Human-Centered Approach to Expanding the Value of AI-Generated Reports in Healthcare. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3706599.3719979

1 Introduction

Healthcare systems across the globe are struggling. As demand for care increases, the available resources and healthcare providers cannot keep up, resulting in rising stress levels and burnout among healthcare staff [7, 14, 31]. On the other hand, patients are frequently experiencing extended delays in receiving care, indicating challenges in the availability of healthcare services. For instance, securing an appointment with a dermatologist can take several months (USA [22] and EU[12, 15]). Artificial Intelligence (AI) has been hoped to alleviate some of those issues [24, 26] through decision support, triage, or automating parts of the healthcare processes [40]. The complexity and variability of support offered within those categories have been increasing after the emergence of Large Language Models (LLMs) and Multi-Modality Models (also referred to as genAI) [36]. Especially, the automated generation of medical reports (also called e-consults and consultant letters) seems to be the answer to many of the troubles of the document-based and ever more asynchronous healthcare [1, 4, 9, 10, 20, 28, 30, 42].

Understanding the significance of medical reports requires a closer look at insights from healthcare research. Such reports are vital tools for communication between specialized clinicians (consultants) and primary care providers (general practitioners, nurse practitioners, or physician assistants). They play a critical role in conveying diagnoses, providing guidance, and supporting patient treatment, which is often described as *informational continuity* [11]. The continuity emphasizes the importance of presenting accurate information at the right time and in a clear manner to ensure effective healthcare delivery [6].

The quality of such reports significantly impacts the ability of primary care providers (PCPs) to deliver appropriate patient care [6, 13, 23]. Conversely, poor communication can lead to miscommunication and inefficiencies, major contributors to medical errors [32]. These errors not only jeopardize patient safety but also result in billions of dollars in avoidable costs that could otherwise be allocated to patient treatment [2].

To mitigate these risks, healthcare researchers have outlined the generally preferred components of consultant reports [27], with a strong preference for structured reporting formats [21]. However, these do not take into consideration concrete medical specialties and the adoption of these recommendations remains inconsistent. Many physicians lack formal training in written communication [8, 19], and the reports they produce may not always align with the needs of their counterparts, such as referring physicians. The communication gap between specialized clinicians and general practitioners is further compounded by differing perspectives on what information is relevant and the timeliness of its delivery, as described by Babington et al. [3].

Computer science research on AI-generated clinical reports has explored some of the communication challenges known to healthcare professionals. Li et al. [17] investigated how LLMs could assist dermatologists by generating initial responses to patient inquiries based on patient inquiries and specialist descriptions of accompanying images. Also supporting patients' ability to understand their health data, Tariq et al. [35] investigate LLMs ability to generate useful supplementary layman's descriptions of chest X-rays. These advancements demonstrate the growing role of LLMs in bridging the gap between medical professionals and patients by making complex clinical information more accessible and understandable.

Another crucial focus of AI research in clinical reporting is ensuring the factual accuracy of AI-generated content – contributing to the *informational continuity*. LLMs are often envisioned as tools to support various aspects of clinical workflows, from data interpretation to diagnosis and communication. This broad applicability introduces potential inaccuracies throughout the process, further exacerbated by the inherent tendency of LLMs to hallucinate [16]. Thus, addressing those risks requires the involvement of specialized clinicians or assessment against clinical ground truth. For example, Shapiro et al. [29] evaluated a chatbot designed to support the full teledermatology consultation process, including generating clinical image descriptions and treatment recommendations. In radiology, Yu et al. [38] proposed RadGraph F1, an automated metric benchmarked against radiologist-assessed report scores. Building on their work, the development teams of MAIRA [30] and its successor MAIRA-2 [5] focused on automating the creation of chest X-ray reports. Bannur and Bouzid et al. [5] further introduced RadFact, an automated evaluation framework to measure the correctness and completeness of these reports. These studies underscore the need for robust evaluation metrics and expert oversight to ensure the factuality of AI-generated clinical content.

While factual correctness is necessary, healthcare research highlights challenges in written medical communication and the dual nature of medical reports - as they serve different purposes to the authors and the recipients. Currently, AI-based systems generating such reports appear to be designed with authors and patients in mind [25]. However, they often overlook the specific needs of recipients, such as PCPs, thereby limiting their potential clinical usefulness. By addressing the information needs of recipients, these reports could increase the efficacy of PCPs and improve patient care. We believe the HCI and CSCW communities are best equipped to inform the design of such AI-based systems, leveraging a deep understanding of how report recipients interact with and utilize these reports.

To contribute to this line of research, improve future genAI-based systems automating clinical text writing, and support responsible AI use in clinical practice, we conducted a co-design study with general practitioners (N=12) from Denmark. This work was part of a larger project between the University of Copenhagen and Melatech investigating the feasibility of LLM reporting in clinical systems. The goal of this study was twofold: (1) *understand the information needs of general practitioners when using a teledermatology service*; and (2) *co-design a dermatology report template that reflects the preferences and informational needs of the recipients, serving as a foundation for AI-generated reports.* Through this study, we make three types of contributions – theoretical, empirical, and practical – to the HCI understanding of genAI use for medical reporting:

Theoretical: A misalignment between the content of specialist-authored medical reports and the information needs of PCPs may hinder PCPs' ability to deliver effective patient care.

Empirical: PCPs expect more information and guidance than they currently receive, however not at the cost of conciseness.

Empirical: The use of smart phrases and templates may induce a "copy-paste" feeling about the reports and reduce PCPs' trust in the recommendations.

Practical: Preferred dermatology report content for rashes and lesions include: *objective visual description*, *assessment*, *primary plan*, *alternative plan*, and only for rashes *follow-up plan.*

2 Study setting

The project this study was a part of was funded by AI Denmark which supports the collaboration of research and business. This collaboration determined the choice of dermatology as the healthcare specialty of focus. Melatech is a start-up offering teledermatology services. Usually, a teledermatology consult aims to provide (a confirmation of) a diagnosis and results in either: (1) confirming a benign condition, (2) treating the patient locally, or (3) referring the patient for treatment at a specialized clinic. In cases when the diagnosis is not possible, a referral is usually advised. The goals of this system are also twofold: (1) reduce the number of unnecessary referrals without jeopardizing patients' health, and (2) offer faster treatments, as waiting times for dermatologist visits in Denmark extend into months.

Topic	Lesions			Rashes		
	f	$L_{K=1}$	$L_{K=2}$	f	$L_{K=1}$	$L_{K=2}$
Subjective (Problem Understanding)	9%	9 (2)	20 (2)	45%	13 (6)	40 (9)
Def: Dermatologist's understanding of what the problem is which the referring PCP needs help with.						
Referral Recap	17%	12 (7)	48 (8)	53%	13 (5)	33 (8)
Def: Dermatologist's understanding of the referral note, clinical history, and patient characteristics.						
Abstract	13%	6 (1)	17 (4)	0%	0 (0)	0 (0)
Def: A short statement capturing the content of the report (high-level positive/negative). Usually located towards the beginning of the report.						
Photo Comments	39%	7 (4)	24 (7)	10%	5 (1)	9 (3)
Def: Dermatologist's evaluation of the quality of the supplied images (good and bad aspects).						
Objective (Visual Description)	85%	17 (6)	39 (14)	91%	13 (5)	33 (10)
Def: Dermatologist's read of the image – a description of objective and visual features.						
Referral Reference	23%	6 (3)	15 (5)	29%	11 (2)	20 (6)
Def: A reference to any questions or suggestions made by the referring PCP in the referral.						
Assessment	82%	17 (9)	52 (27)	95%	18 (9)	45 (11)
Def: Dermatologist's interpretation of the visual findings in the light of the referral and the problem understanding.						
Education	17%	22 (8)	73 (23)	21%	17 (7)	47 (18)
Def: Definitory content (incl. links to additional resources) that may help the referring PCP to understand the assessment and/or treatment plan.						
Primary Plan	88%	13 (8)	44 (20)	100%	28 (10)	60 (20)
Def: Dermatologist's suggestion for immediate next steps (treatment, patient care, dismissal).						
Alternative Plan	15%	16 (10)	76 (1)	11 %	28 (8)	81 (37)
Def: An alternative course of action to the first suggestion, either because there are possible alternative treatment plans or because of the conditionality of the assessment.						
Follow-up Plan	15%	16 (5)	40 (9)	50%	21 (8)	71 (15)
Def: Dermatologist's suggestion for the next steps after the completion of the primary or the alternative plans.						
Comment	26%	20 (15)	102 (28)	11%	20 (8)	73 (9)
Def: Any other content that may be relevant for the PCP or the patient to know in regard to this assessment.						

Table 1: Twelve topics observed in the analyzed dermatology reports including their definitions (conceptualized by the first author), frequency of use in the reports, two centroids of K=2 clustering of topics by word count (SD).

3 Method

To inform our co-design study of a dermatology report template that reflects the preferences and informational needs of the recipients, we conducted extensive document analysis of real dermatology reports.

3.1 Informing the co-design study

In collaboration with Melatech, we acquired 162 anonymized dermatology reports treating rashes (80) and lesions (82) from Denmark and the US. From July to August 2024, the first author line-by-line coded all the reports (in their original languages) describing each excerpt based on the content it described and its position in the report. Each of the excerpts was assigned to a suitable topic. The initial topics were informed by the SOAP (subjective, objective, assessment, plan) reporting guidelines [18], however, they almost immediately became too limiting, as dermatologists covered a wider array of topics in their reports. When no topic was suitable, a new one was added and the already assigned codes were adjusted. After analysis completion, twelve topics were established and validated by the second author - a medical doctor (Tab. 1).

Throughout this analysis, we observed that reports varied in length and that the topics discussed depended on the condition (lesions or rashes) described. To further explore these variations, we conducted a separate analysis for each condition. Given the varying word counts across topics, we applied $K = 2$ clustering to identify two centroids that better represented the observed word count distribution than a simple median. Additionally, we calculated the frequency of each topic's use. This approach allowed us to preserve the variability within the data while generalizing it sufficiently to inform the design of a report template. The results of this analysis are presented in table 1.

Importantly, this analysis was conducted to inform the co-design study, rather than serving as the sole foundation for the report templates. We acknowledge the limitations of our approach, including annotations performed by a single person and the use of clustering on topics with low frequency. However, the insights derived from this analysis were instrumental in shaping the co-design study, where participating medical doctors critically evaluated and refined them.

3.2 Designing the co-design study

Having gained an understanding of the dermatology report structure, we proceeded to design the co-design study. In collaboration with Melatech, we selected six representative cases of rashes (1 image each) and lesions (2 images each). We strove to include one easy and one difficult case, and one case with sub-par image quality. Then, the second author proposed plausible clinical referrals for those patients. Using the images, the referrals, and a guide comprising the topics, examples, and desired word counts including standard deviations, the second author wrote possible content for each of the topics for each of the cases aiming to match the desired word count, we will call them topic snippets. Not all of the topics were used for all the cases. While we aimed to exhaust the offered information, not all the topics made sense in the context of every case. The decision was left to the medical expertise of the second author.

In the co-design sessions, we used the topic definitions, topic snippets, images, and referral notes. We prepared two versions of the study: in-person and online. For the in-person version, we printed all of the materials, including two copies of the topic snippets. One of the copies was kept as is, and the other one was cut so that each of the snippets was on a separate piece of paper. This approach provided an opportunity to get an easy overview of the snippets, while individual snippets were used to assemble the reports. For the online version, the content was moved to a digital whiteboard - Miro[1].

3.3 Conducting the co-design study

The sessions comprised two parts: (1) a short interview and (2) co-design of the preferred dermatology reports for the prepared cases. The interview focused on understanding what makes dermatology reports useful to PCPs, what makes a report sub-par, and what makes a report not useful. Following the interview, the participants were asked to familiarize themselves with the twelve topics. Next, the process was the same for each case. Participants were asked to study the provided images and the referral note, imagining it as a case they had sent for a teledermatology consultation. They then assembled their preferred report for this consultation from the prepared snippets. Once the report was complete, we asked questions about its content: why they chose specific snippets in the given lengths, why they omitted others, and whether any important information was missing from the prepared snippets.

The co-design studies were conducted between September 1 and November 1, 2024. We recruited 12 specialists in general practice (11+ years of experience) from Denmark who worked in primary care clinics and had experience with teledermatology (10 using Melatech's system). Eight of the sessions were conducted in person, and four online. In total, 72 reports were assembled. The sessions took approximately 45 minutes each and were conducted by the first author. The collected data included notes and pictures of the obtained reports. No audio or other personal or patient data were collected during the sessions, oral consent was obtained for the pictures. The first author analyzed the composition of assembled reports and the length of the topics used, considering cases and conditions. Next, the first author compared the notes taken line-by-line against the quantitative results to explore possible explanations and to foreground any alignments and discrepancies between these two data types.

[1]miro.com

4 Findings

4.1 The dermatology report template

Based on the analysis of the composition of the co-designed reports, we can see the emergence of five foundational topics: *objective visual description*, *assessment*, *primary plan*, *alternative plan*, and only for rashes *follow-up plan* (Fig. 2).

These results contrast with the widely disseminated SOAP structure. First, our participants put less emphasis on (reflected in low usage of) topics addressing clinical information, as it was always available in their electronic health record system and was considered superfluous in a dermatology report. Second, significantly more emphasis was placed on the three plan topics. While also present in SOAP, only the *primary plan* was consistently observed in the real-world reports (Tab. 1). Consistent selection of those snippets suggests that providing exhaustive treatment options and supporting PCPs in making informed decisions about the treatment course was considered useful. The low usage of *follow-up plan* in lesions was related to the nature of those conditions, which either warranted excision or were benign and could be dismissed. Thus, PCPs had a good understanding of the next steps.

The *comments* and *education* snippets elicited mixed reactions from our participants. Some PCPs appreciated these sections as helpful reminders of relevant clinical information that might be overlooked in the heat of daily practice. Others, however, found them unnecessary, viewing them as almost patronizing. It is worth mentioning that these snippets mostly provided general guidance and reference information about the conditions at hand - sometimes in the form of a link, which was almost never preferred. Our participants reflected that during practice, switching contexts and reading up on certain conditions seemed highly unlikely.

When it came to the preferred length of snippets, the general consensus was clear: the more concise, the better (Fig. 3). However, this preference came with a critical caveat: the comprehensiveness of the information within the snippet often outweighed the need for brevity. PCPs consistently favored snippets that provided detailed information, particularly when it helped them better understand and manage a case. This preference was evident in the selection of longer snippets for topics such as *objective (visual description)* and *assessment*. This preference was particularly prominent in the *primary plan* topic in case 3, where participants overwhelmingly opted for the longer snippet. The two options for case 3 read as follows: (short) *should be excised and sent for histology;* (long) *if you are comfortable with it, I recommend excision with a few mm margins and direct closure under local anesthetic. Closure should be carried out in the longitudinal direction of the arm (proximal → distal) to minimize pulling on the wound.* Nine out of twelve participants selected the longer version. This choice was especially consequential, as it directly influenced their treatment approach. Participants explained that receiving the shorter version would have led them to refer the patient to a specialist for excision, even if they could perform the

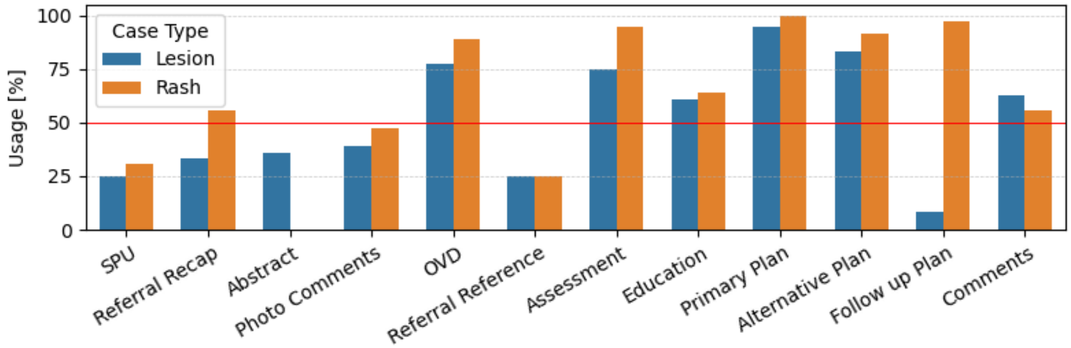

Figure 2: Aggregated usage of topics for all the participants across two types of investigated conditions. SPU - Subjective (Problem Understanding), OVD - Objective (Visual Description)

procedure themselves. Conversely, if they received detailed instructions in the longer snippet, they would perform the procedure in their clinic, reducing patient waiting time, minimizing risks, and alleviating pressure on the healthcare system. This example underscores how seemingly simple decisions about report formatting and content can have far-reaching implications highlighting the critical balance between brevity and comprehensiveness in meeting the information needs of PCPs.

4.2 Meeting the Information Needs of PCPs Through Dermatology Reports

Every participant reported that the conceptualized topics covered their information needs, with no expressed wish for additional content. In fact, most of the assembled reports did not use all of the available topics, with the shortest ones comprising only two snippets (*assessment* and *primary plan*).

The quality of dermatology reports in the real world often does not support PCPs' provision of the best care possible (N=10). Our participants found some of the reports, encountered in practice, insubstantial in guidance and rationale. As the guidance aspect of the reports was explored in the section above, we will focus on the rationale. Many of the PCPs considered teledermatology consults an opportunity for learning and self-improvement, the goal being sending fewer referrals and being able to take care of more patients internally. To do that, they used the information included in the *objective (visual description)* topic. An in-depth grounding of dermatology assessment in the visual features was considered crucial for their learning, hence the preference for longer snippets.

Finally, our participants reflected on the tone of the reports, they appreciated the to-the-point, personal, and "human" qualities of the prepared snippets. Some of our participants compared them against reports seen in their practice that seemed like pre-defined templates populated with smart phrases, a common occurrence in medical reports. However, one PCP specifically noted that if a report appears overly copy-pasted, it undermines their confidence in the assessment, as they feel the consulting dermatologist did not carefully consider their patient's case but merely selected a template and sent the report. This observation suggests it is not only the content of the reports that matters but also their format that influences the perceived usefulness of dermatology reports.

5 Discussion And Future Work

LLMs have been increasingly investigated to support clinical reporting [30, 33, 34, 36] and although referring physicians are recognized as the main recipients and beneficiaries of those reports, to our best knowledge, their perspective has not been taken into consideration when investigating genAI reporting [25, 36]. We decided to explore the previously omitted perspective of report recipients. This study showed that reports written by consultant clinicians may not always meet the information and guidance needs of report recipients. We recognized this as an opportunity for HCI researchers to investigate how AI technologies can bridge that gap, leading to improved patient outcomes and reduced strain on the healthcare system. To start this line of research, we completed the first step of engaging report recipients to understand their information needs and co-design report templates that can be used to inform AI generation of reports in the future. The results of this study confirmed the misalignment of expectations for the content of consultant reports. We will use this groundwork to conduct three more studies to better understand how genAI-based systems can realistically improve clinical reporting.

First, we acknowledge the limitation of our study in focusing solely on PCPs from Denmark when exploring the report structure. The information gap discovered may be related to the incentive structure of public healthcare systems and should be further investigated. Moreover, the proposed structure reflects PCPs' preferences. However, it may inadvertently require more work from dermatologists. Even limiting the overhead to providing additional input for LLMs to generate reports—an outcome that may not be practical in real-world settings. To address this, we are recruiting dermatologists to evaluate the PCP-designed report structure and understand their rationale when composing reports, ensuring the design aligns with both perspectives. We hope this study will contribute to the broader body of HCI work on designing systems that balance the diverse needs of specialized and primary care providers, fostering effective collaboration and knowledge transfer.

Second, responding to the HCI call for greater inclusion of patients in research on clinical AI-based systems [37], we plan to conduct a co-design study to provide a third perspective on clinical reports to complete the picture. Existing research has explored how LLMs might generate patient-friendly reports [17, 35], but these

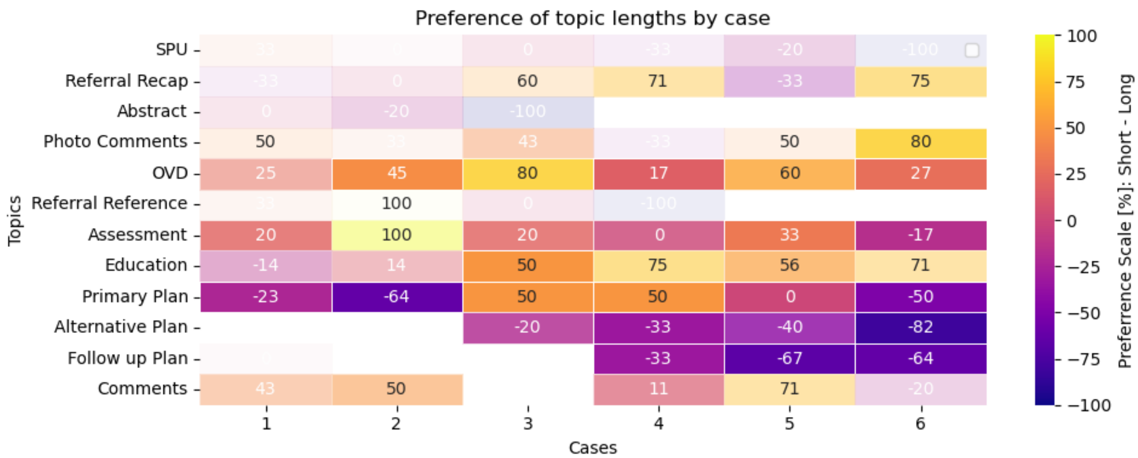

Figure 3: Aggregated usage of topics for all the participants across cases. The more transparent a cell is, the fewer participants chose that topic. The more yellow, the higher the consensus for long descriptions. The more purple, the higher the consensus for short descriptions. SPU - Subjective (Problem Understanding), OVD - Objective (Visual Description)

studies often focus on access to medical data outside regular clinical interactions, which are typically mediated and supported by patients' PCPs. We see this as an opportunity to investigate the utility of AI-generated reports at the intersection of patients, general practitioners, and specialists, where collaborative decision-making occurs. This approach can help identify how LLMs may enhance understanding and communication among these key stakeholders.

Finally, aware of the uniqueness of local clinical practices, we believe the HCI community should lead research into new ways of envisioning, configuring, and interacting with AI in safety-critical environments [39, 41]. This aligns with prior HCI work emphasizing "the need for a broadened focus beyond 'chatbots' and 'AI agents' as concepts for how we envision AI to benefit clinical workflows" [36]. To contribute to this line of work and build on the results of this study, over the past months, we have been iteratively developing a prototype of an LLM-based reporting system that uses our co-designed report templates. This work will continue, as we are collaborating with dermatologists from Denmark, Sweden, and the US to investigate two issues: (1) assessing through simulated workflows the viability of using a PCP-derived template to generate medical reports; and (2) exploring the design space of interaction techniques with LLMs for text generation in safety-critical environments. We aim for our future studies to provide the HCI community with actionable insights into designing LLM-based systems that do not rely on chat-based interactions, enhance knowledge transfer, and support collaborative clinical work.

Acknowledgments

We would like to thank our participants for their time and valuable insights. The icons in Figure 1 were designed by Kiranshastry (report), DinosoftLabs (doctor), Freepik (patient, dermatologist, rash), msidiqf (phone camera) from flaticon.com

References

[1] Sri Likhita Adru and S Srijayanthi. 2024. Integrated AI-Driven Healthcare System: Predictive Diagnosis, Personalized Medical-Report Content Generation, and Physician Recommendation. In *2024 2nd International Conference on Sustainable Computing and Smart Systems (ICSCSS)*. 1148–1153. https://doi.org/10.1109/ICSCSS60660.2024.10625283

[2] Ritu Agarwal, Daniel Z. Sands, and Jorge Díaz Schneider. 2010. Quantifying the Economic Impact of Communication Inefficiencies in U.S. Hospitals. *Journal of Healthcare Management* 55, 4 (Aug. 2010), 265. https://journals.lww.com/jhmonline/abstract/2010/07000/Quantifying_the_Economic_Impact_of_Communication.7.aspx/1000

[3] S Babington, C Wynne, Ch Atkinson, Be Hickey, and As Abdelaal. 2003. Oncology service correspondence: Do we communicate? *Australasian Radiology* 47, 1 (2003), 50–54. https://doi.org/10.1046/j.1440-1673.2003.01129.x _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1046/j.1440-1673.2003.01129.x.

[4] Richard S. Bakalar. 2022. Telemedicine: Its Past, Present and Future. In *Healthcare Information Management Systems: Cases, Strategies, and Solutions*, Joan M. Kiel, George R. Kim, and Marion J. Ball (Eds.). Springer International Publishing, Cham, 149–160. https://doi.org/10.1007/978-3-031-07912-2_9

[5] Shruthi Bannur, Kenza Bouzid, Daniel C. Castro, Anton Schwaighofer, Anja Thieme, Sam Bond-Taylor, Maximilian Ilse, Fernando Pérez-García, Valentina Salvatelli, Harshita Sharma, Felix Meissen, Mercy Ranjit, Shaury Srivastav, Julia Gong, Noel C. F. Codella, Fabian Falck, Ozan Oktay, Matthew P. Lungren, Maria Teodora Wetscherek, Javier Alvarez-Valle, and Stephanie L. Hyland. 2024. MAIRA-2: Grounded Radiology Report Generation. https://doi.org/10.48550/arXiv.2406.04449 arXiv:2406.04449 [cs].

[6] Whitney Berta, Jan Barnsley, Jeff Bloom, Rhonda Cockerill, Dave Davis, Liisa Jaakkimainen, Anne Marie Mior, Yves Talbot, and Eugene Vayda. 2009. Enhancing continuity of information: Essential components of consultation reports. *Canadian Family Physician* 55, 6 (June 2009), 624–625.e5. https://www.cfp.ca/content/55/6/624 Publisher: The College of Family Physicians of Canada Section: Research.

[7] Diana Duong and Lauren Vogel. 2023. Overworked health workers are "past the point of exhaustion". *Canadian Medical Association Journal* 195, 8 (Feb. 2023), E309–E310. https://doi.org/10.1503/cmaj.1096042

[8] José François. 2011. Tool to assess the quality of consultation and referral request letters in family medicine. *Canadian Family Physician* 57, 5 (May 2011), 574–575. https://www.cfp.ca/content/57/5/574 Publisher: The College of Family Physicians of Canada Section: Practice.

[9] Andrea Giacalone, Luca Marin, Massimiliano Febbi, Thomas Franchi, and Marcos Roberto Tovani-Palone. 2022. eHealth, telehealth, and telemedicine in the management of the COVID-19 pandemic and beyond: Lessons learned and future perspectives. *World Journal of Clinical Cases* 10, 8 (March 2022), 2363–2368. https://doi.org/10.12998/wjcc.v10.i8.2363

[10] Katherine E. Goodman, Paul H. Yi, and Daniel J. Morgan. 2024. AI-Generated Clinical Summaries Require More Than Accuracy. *JAMA* 331, 8 (Feb. 2024), 637–638. https://doi.org/10.1001/jama.2024.0555

[11] Jeannie L. Haggerty, Robert J. Reid, George K. Freeman, Barbara H. Starfield, Carol E. Adair, and Rachael McKendry. 2003. Continuity of care: a multidisciplinary review. *BMJ* 327, 7425 (Nov. 2003), 1219–1221. https://doi.org/10.1136/bmj.327.7425.1219 Publisher: British Medical Journal Publishing Group Section: Education and debate.

[12] M. Hald, T. Agner, J. Blands, J.D. Johansen, and on behalf of the Danish Contact Dermatitis Group. 2009. Delay in medical attention to hand eczema: a follow-up study. *British Journal of Dermatology* 161, 6 (Dec. 2009), 1294–1300. https://doi.org/10.1111/j.1365-2133.2009.09402.x

[13] Ragnhild Hellesø, Margarethe Lorensen, and Lena Sorensen. 2004. Challenging the information gap – the patients transfer from hospital to home health care. *International Journal of Medical Informatics* 73, 7 (Aug. 2004), 569–580. https://doi.org/10.1016/j.ijmedinf.2004.04.009

[14] Yinhuan Hu and Zixia Zhang. 2015. Skilled doctors in tertiary hospitals are already overworked in China. *The Lancet Global Health* 3, 12 (Dec. 2015), e737. https://doi.org/10.1016/S2214-109X(15)00192-8 Publisher: Elsevier.

[15] François Kraus and Hugo Lasserre. [n. d.]. *Renonciation aux soins, complexes, stigmates... LES FRANÇAIS FACE AUX MALADIES DE PEAU ET A L'ECZEMA : LA GRANDE ENQUETE.* Technical Report. https://www.ifop.com/publication/renonciation-aux-soins-complexes-stigmates-les-francais-face-aux-maladies-de-peau-et-a-leczema-la-grande-enquete/

[16] Florian Leiser, Sven Eckhardt, Valentin Leuthe, Merlin Knaeble, Alexander Mädche, Gerhard Schwabe, and Ali Sunyaev. 2024. HILL: A Hallucination Identifier for Large Language Models. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3613904.3642428

[17] Brenna Li, Amy Wang, Patricia Strachan, Julie Anne Séguin, Sami Lachgar, Karyn C Schroeder, Mathias S Fleck, Renee Wong, Alan Karthikesalingam, Vivek Natarajan, Yossi Matias, Greg S Corrado, Dale Webster, Yun Liu, Naama Hammel, Rory Sayres, Christopher Semturs, and Mike Schaekermann. 2024. Conversational AI in health: Design considerations from a Wizard-of-Oz dermatology case study with users, clinicians and a medical LLM. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–10. https://doi.org/10.1145/3613905.3651891

[18] Chen-Tan Lin, Marlene McKenzie, Jonathan Pell, and Liron Caplan. 2013. Health Care Provider Satisfaction With a New Electronic Progress Note Format: SOAP vs APSO Format. *JAMA Internal Medicine* 173, 2 (Jan. 2013), 160–162. https://doi.org/10.1001/2013.jamainternmed.474

[19] Lorelei Lingard, Brian Hodges, Helen Macrae, and Risa Freeman. 2004. Expert and trainee determinations of rhetorical relevance in referral and consultation letters. *Medical Education* 38, 2 (2004), 168–176. https://doi.org/10.1111/j.1365-2923.2004.01745.x _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2923.2004.01745.x.

[20] Guanxiong Liu, Tzu-Ming Harry Hsu, Matthew McDermott, Willie Boag, Wei-Hung Weng, Peter Szolovits, and Marzyeh Ghassemi. 2019. Clinically Accurate Chest X-Ray Report Generation. In *Proceedings of the 4th Machine Learning for Healthcare Conference*. PMLR, 249–269. https://proceedings.mlr.press/v106/liu19a.html ISSN: 2640-3498.

[21] C. Melville, S. Hands, and P. Jones. 2002. Randomised trial of the effects of structuring clinic correspondence. *Archives of Disease in Childhood* 86, 5 (May 2002), 374–375. https://doi.org/10.1136/adc.86.5.374 Publisher: BMJ Publishing Group Ltd Section: Acute paediatrics.

[22] Society of Dermatology Physician Associates. [n. d.]. *Patients Are Waiting: America's Dermatology Wait Times Crisis.* Technical Report. Society of Dermatology Physician Associates. 20 pages. https://www.dermpa.org/page/GAPP

[23] Ann S. O'Malley and James D. Reschovsky. 2011. Referral and Consultation Communication Between Primary Care and Specialist Physicians: Finding Common Ground. *Archives of Internal Medicine* 171, 1 (Jan. 2011). https://doi.org/10.1001/archinternmed.2010.480

[24] Trishan Panch, Heather Mattie, and Leo Anthony Celi. 2019. The "inconvenient truth" about AI in healthcare. *npj Digital Medicine* 2, 1 (Aug. 2019), 1–3. https://doi.org/10.1038/s41746-019-0155-4 Publisher: Nature Publishing Group.

[25] Jiwoo Park, Kangrok Oh, Kyunghwa Han, and Young Han Lee. 2024. Patient-centered radiology reports with generative artificial intelligence: adding value to radiology reporting. *Scientific Reports* 14, 1 (June 2024), 13218. https://doi.org/10.1038/s41598-024-63824-z

[26] Pranav Rajpurkar, Emma Chen, Oishi Banerjee, and Eric J. Topol. 2022. AI in health and medicine. *Nature Medicine* 28, 1 (Jan. 2022), 31–38. https://doi.org/10.1038/s41591-021-01614-0 Publisher: Nature Publishing Group.

[27] Arjun H. Rash, Robert Sheldon, Maoliosa Donald, Cindy Eronmwon, and Vikas P. Kuriachan. 2018. Valued Components of a Consultant Letter from Referring Physicians' Perspective: a Systematic Literature Synthesis. *Journal of General Internal Medicine* 33, 6 (June 2018), 948–954. https://doi.org/10.1007/s11606-018-4356-3

[28] Marwareed Rehman, Imran Shafi, Jamil Ahmad, Carlos Osorio Garcia, Alina Eugenia Pascual Barrera, and Imran Ashraf. 2024. Advancement in medical report generation: current practices, challenges, and future directions. *Medical & Biological Engineering & Computing* (Dec. 2024). https://doi.org/10.1007/s11517-024-03265-y

[29] Jonathan Shapiro and Anna Lyakhovitsky. 2024. Revolutionizing teledermatology: Exploring the integration of artificial intelligence, including Generative Pre-trained Transformer chatbots for artificial intelligence-driven anamnesis, diagnosis, and treatment plans. *Clinics in Dermatology* 42, 5 (Sept. 2024), 492–497. https://doi.org/10.1016/j.clindermatol.2024.06.020

[30] Shaury Srivastav, Mercy Ranjit, Fernando Pérez-García, Kenza Bouzid, Shruthi Bannur, Daniel C. Castro, Anton Schwaighofer, Harshita Sharma, Maximilian Ilse, Valentina Salvatelli, Sam Bond-Taylor, Fabian Falck, Anja Thieme, Hannah Richardson, Matthew P. Lungren, Stephanie L. Hyland, and Javier Alvarez-Valle. 2024. MAIRA at RRG24: A specialised large multimodal model for radiology report generation. In *Proceedings of the 23rd Workshop on Biomedical Natural Language Processing*, Dina Demner-Fushman, Sophia Ananiadou, Makoto Miwa, Kirk Roberts, and Junichi Tsujii (Eds.). Association for Computational Linguistics, Bangkok, Thailand, 597–602. https://doi.org/10.18653/v1/2024.bionlp-1.50

[31] Melissa Suran. 2023. Overworked and Understaffed, More Than 1 in 4 US Nurses Say They Plan to Leave the Profession. *JAMA* 330, 16 (Oct. 2023), 1512–1514. https://doi.org/10.1001/jama.2023.10055

[32] Kathleen M. Sutcliffe, Elizabeth Lewton, and Marilynn M. Rosenthal. 2004. Communication Failures: An Insidious Contributor to Medical Mishaps. *Academic Medicine* 79, 2 (Feb. 2004), 186. https://journals.lww.com/academicmedicine/fulltext/2004/02000/Communication_Failures__An_Insidious_Contributor.19.aspx

[33] Ming Tai-Seale, Sally L. Baxter, Florin Vaida, Amanda Walker, Amy M. Sitapati, Chad Osborne, Joseph Diaz, Nimit Desai, Sophie Webb, Gregory Polston, Teresa Helsten, Erin Gross, Jessica Thackaberry, Ammar Mandvi, Dustin Lillie, Steve Li, Geneen Gin, Suraj Achar, Heather Hofflich, Christopher Sharp, Marlene Millen, and Christopher A. Longhurst. 2024. AI-Generated Draft Replies Integrated Into Health Records and Physicians' Electronic Communication. *JAMA Network Open* 7, 4 (April 2024), e246565. https://doi.org/10.1001/jamanetworkopen.2024.6565

[34] Ryutaro Tanno, David G. T. Barrett, Andrew Sellergren, Sumedh Ghaisas, Sumanth Dathathri, Abigail See, Johannes Welbl, Charles Lau, Tao Tu, Shekoofeh Azizi, Karan Singhal, Mike Schaekermann, Rhys May, Roy Lee, SiWai Man, Sara Mahdavi, Zahra Ahmed, Yossi Matias, Joelle Barral, S. M. Ali Eslami, Danielle Belgrave, Yun Liu, Sreenivasa Raju Kalidindi, Shravya Shetty, Vivek Natarajan, Pushmeet Kohli, Po-Sen Huang, Alan Karthikesalingam, and Ira Ktena. 2024. Collaboration between clinicians and vision–language models in radiology report generation. *Nature Medicine* (Nov. 2024), 1–10. https://doi.org/10.1038/s41591-024-03302-1 Publisher: Nature Publishing Group.

[35] Amara Tariq, Shubham Trivedi, Aisha Urooj, Gokul Ramasamy, Sam Fathizadeh, Matthew Stib, Nelly Tan, Bhavik Patel, and Imon Banerjee. 2024. Patient-centric Summarization of Radiology Findings using Two-step Training of Large Language Models. *ACM Trans. Comput. Healthcare* (Dec. 2024). https://doi.org/10.1145/3709154 Just Accepted.

[36] Nur Yildirim, Hannah Richardson, Maria Teodora Wetscherek, Junaid Bajwa, Joseph Jacob, Mark Ames Pinnock, Stephen Harris, Daniel Coelho De Castro, Shruthi Bannur, Stephanie Hyland, Pratik Ghosh, Mercy Ranjit, Kenza Bouzid, Anton Schwaighofer, Fernando Pérez-García, Harshita Sharma, Ozan Oktay, Matthew Lungren, Javier Alvarez-Valle, Aditya Nori, and Anja Thieme. 2024. Multimodal Healthcare AI: Identifying and Designing Clinically Relevant Vision-Language Applications for Radiology. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–22. https://doi.org/10.1145/3613904.3642013

[37] Dong Whi Yoo, Hayoung Woo, Viet Cuong Nguyen, Michael L. Birnbaum, Kaylee Payne Kruzan, Jennifer G Kim, Gregory D. Abowd, and Munmun De Choudhury. 2024. Patient Perspectives on AI-Driven Predictions of Schizophrenia Relapses: Understanding Concerns and Opportunities for Self-Care and Treatment. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–20. https://doi.org/10.1145/3613904.3642369

[38] Feiyang Yu, Mark Endo, Rayan Krishnan, Ian Pan, Andy Tsai, Eduardo Pontes Reis, Eduardo Kaiser Ururahy Nunes Fonseca, Henrique Min Ho Lee, Zahra Shakeri Hossein Abad, Andrew Y. Ng, Curtis P. Langlotz, Vasantha Kumar Venugopal, and Pranav Rajpurkar. 2023. Evaluating progress in automatic chest X-ray radiology report generation. *Patterns* 4, 9 (Sept. 2023). https://doi.org/10.1016/j.patter.2023.100802 Publisher: Elsevier.

[39] Hubert D. Zając, Tariq O. Andersen, Elijah Kwasa, Ruth Wanjohi, Mary K. Onyinkwa, Edward K. Mwaniki, Samuel N. Gitau, Shawnim S. Yaseen, Jonathan F. Carlsen, Marco Fraccaro, Michael B. Nielsen, and Yunan Chen. 2025. Towards Clinically Useful AI: From Radiology Practices in Global South and North to Visions of AI Support. *ACM Trans. Comput.-Hum. Interact.* (Feb. 2025). https://doi.org/10.1145/3715115 Just Accepted.

[40] Hubert D. Zając, Dana Li, Xiang Dai, Jonathan F. Carlsen, Finn Kensing, and Tariq O. Andersen. 2023. Clinician-Facing AI in the Wild: Taking Stock of the Sociotechnical Challenges and Opportunities for HCI. *ACM Trans. Comput.-Hum. Interact.* 30, 2 (March 2023), 33:1–33:39. https://doi.org/10.1145/3582430

[41] Hubert Dariusz Zając, Jorge Miguel Neves Ribeiro, Silvia Ingala, Simona Gentile, Ruth Wanjohi, Samuel Nguku Gitau, Jonathan Frederik Carlsen, Michael Bachmann Nielsen, and Tariq Osman Andersen. 2024. "It depends": Configuring AI to Improve Clinical Usefulness Across Contexts. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference (DIS '24)*. Association for Computing Machinery, New York, NY, USA, 874–889. https://doi.org/10.1145/3643834.3660707

[42] Jonah Zaretsky, Jeong Min Kim, Samuel Baskharoun, Yunan Zhao, Jonathan Austrian, Yindalon Aphinyanaphongs, Ravi Gupta, Saul B. Blecker, and Jonah Feldman. 2024. Generative Artificial Intelligence to Transform Inpatient Discharge Summaries to Patient-Friendly Language and Format. *JAMA Network Open* 7, 3 (March 2024), e240357. https://doi.org/10.1001/jamanetworkopen.2024.0357