Rapport Matters: Enhancing HIV mHealth Communication through Linguistic Analysis and Large Language Models

Zhiyuan Wang*
vmf9pr@virginia.edu
Department of Systems and Information Engineering, University of Virginia
Charlottesville, Virginia, USA

Varun Reddy
dpc3qt@virginia.edu
Department of Computer Science, University of Virginia
Charlottesville, Virginia, USA

Karen Ingersoll
kes7a@uvahealth.org
Department of Psychiatry and Neurobehavioral Sciences, University of Virginia
Charlottesville, Virginia, USA

Tabor Flickinger
tes3j@virginia.edu
School of Medicine, University of Virginia
Charlottesville, Virginia, USA

Laura E. Barnes
lb2dp@virginia.edu
Department of Systems and Information Engineering, University of Virginia
Charlottesville, Virginia, USA

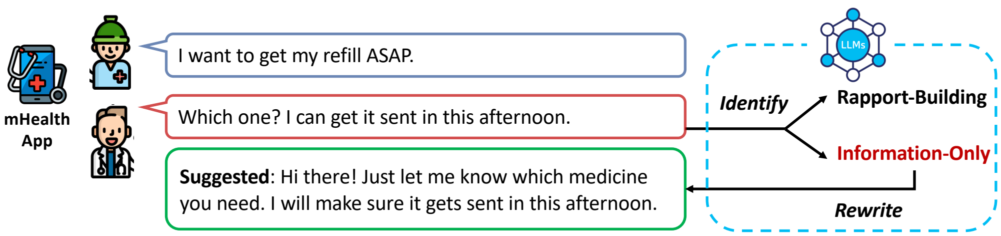

Figure 1: An Example of AI-Enhanced Rapport Building in HIV mHealth Conversations.

ABSTRACT

In HIV care, a strong rapport between patient and provider is essential for strengthening trust, enhancing therapy adherence, and ultimately leading to improved health outcomes. As the adoption of digital interactions in HIV care via mobile health (mHealth) tools is emerging, maintaining rapport in these asynchronous text-based communications becomes a critical yet challenging task. In this paper, we analyze 1,740 messages from an mHealth platform, categorized by experienced clinicians as either 'rapport-building' or 'information-only.' We utilize linguistic analysis to uncover key attributes of rapport-building communication. This led to a set of machine learning (ML) models and Large Language Models (LLMs) capable of classifying these communication styles. Further, we propose the application of LLMs not only to identify but also to actively rewrite 'information only' messages into versions that enhance rapport building without compromising information integrity. Our research demonstrates potential advancements in HIV mHealth communication by integrating linguistic analysis with language models, leading to more effective patient-provider interactions.

*Corresponding author. Email: vmf9pr@virginia.edu

CHI EA '24, May 11–16, 2024, Honolulu, HI, USA

ACM ISBN 979-8-4007-0331-7/24/05
https://doi.org/10.1145/3613905.3651077

CCS CONCEPTS

• **Applied computing** → **Health care information systems**; • **Computing methodologies** → **Natural language processing**; • **Human-centered computing** → *Collaborative and social computing*.

KEYWORDS

Patient-clinician communication, mobile health, large language models

ACM Reference Format:
Zhiyuan Wang, Varun Reddy, Karen Ingersoll, Tabor Flickinger, and Laura E. Barnes. 2024. Rapport Matters: Enhancing HIV mHealth Communication through Linguistic Analysis and Large Language Models. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '24), May 11–16, 2024, Honolulu, HI, USA.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3613905.3651077

1 INTRODUCTION

The integration of mobile health (mHealth) technologies in Human Immunodeficiency Virus (HIV) care has not only revolutionized patient-provider communication but also has profound implications for patients' overall health outcomes and quality of life [7, 11, 14].

This advancement is crucial in HIV care, where effective interpersonal interaction extends **beyond information exchange to building rapport** and addressing psychological aspects of living with HIV [22, 33]. It also plays a vital role in reducing medical errors, improving symptom management, and increasing patient adherence [4, 5, 20, 54]. Additionally, addressing social determinants and societal disparities like stigma through accessible communication is particularly key to improving the well-being among people living with HIV [8, 22].

Clinicians' ability to respond to patients' emotional cues impacts patient satisfaction and clinical outcomes [25], including in HIV care [19]. Despite these benefits, studies of patient-clinician communication have revealed frequent missed opportunities with patient cues that are not detected or not responded to in a therapeutic manner [17, 34]. Missed opportunities tend to be more frequent for patients who may already be disadvantaged or stigmatized in the healthcare system, further exacerbating health disparities [36]. Digital communication can be relationship-building and patient-centered, but gaps exist in achieving this goal in clinical practice [24]. Clinicians perceive the value of secure messaging in enhancing patient care, but they also report challenges in its use, including time constraints, difficulty interpreting and managing messages, and concerns about how best to use messaging to build relationships [3]. Therefore, there is a need to develop tools to improve communication quality in messaging.

The subtleties of rapport-building in digital interactions are often lost or overlooked. Despite progress in communication skills training for healthcare providers, specific training and best practices for digital, text-based encounters remain underdeveloped [6, 50]. Recent advances in AI and interactive techniques hold great promise to bridge this gap, with computational methods providing novel opportunities to improve healthcare communication quality [21, 38, 45]. Our study aims to fill this gap by exploring the intersection of AI, linguistics, and healthcare communication in the context of HIV care, aiming to develop computational methods that assess and optimize rapport building in text-based digital communication [46].

This paper focuses on understanding and enhancing rapport in HIV mHealth conversations using linguistic analysis and LLM-driven tools. As depicted in Figure 1, we envision a system that not only identifies but also enhances the quality of message exchanges, rewriting information-only messages into more rapport-building communications. This approach exemplifies the practical application of our research and forms our primary research questions:

RQ1: What are the linguistic attributes of 'rapport-building' messages compared to the ones only delivering information in HIV mHealth conversations?

RQ2: Can we **identify** providers' 'rapport-building' and 'information-only' messages in HIV mHealth conversations using natural language processing (NLP) methods?

RQ3: Can we enhance the rapport by **rewriting** providers' 'information-only' messages to 'rapport-building' ones without losing information integrity?

This paper introduces a three-phase study leveraging a pilot dataset from a mHealth HIV care digital platform named PositiveLinks. 1,740 messages exchanged by patients, providers, and staff based at a Ryan White HIV Clinic. Messages were sent and received through the app, which includes secure messaging as one feature of a multi-component application [11]. Licensed clinicians have categorized these messages as either 'rapport-building' or 'information-only' [18]. Our approach starts with a linguistic analysis to identify key indicators of rapport-building (RQ1), followed by the development of classification models (RQ2), and culminates in the creation of an LLM-based tool for message enhancement (RQ3). Each phase is designed to progressively build upon the previous, illustrating a comprehensive investigation from theoretical analysis to practical application in enhancing HIV mHealth communication. The contributions of this paper include:

- We conducted a linguistic analysis to identify characteristics of rapport-building in HIV mHealth communications using a pilot dataset.
- We developed and validated ML and LLM models for classifying rapport-building messages, enhancing AI's role in patient-provider communication.
- We proposed a method using LLMs to transform 'information-only' messages into 'rapport-building' ones, maintaining information integrity while enhancing rapport.
- Combining together, this work provides a novel, scalable method for mHealth platforms to enhance patient-provider communication, adaptable to healthcare scenarios beyond HIV care.

2 RELATED WORK

This paper builds upon the growing body of research in mHealth, highlighting the importance of effective patient-provider communication for improved health outcomes particularly in HIV care [11, 30, 35, 51]. We draw inspiration from studies demonstrating the efficacy of mHealth platforms like PositiveLinks in enhancing care engagement [11].

In the HCI and CSCW domains, there is a growing focus on enhancing patient-provider communication. Studies have explored various aspects, such as the effectiveness of online platforms in facilitating patient-doctor interactions [16], addressing communication barriers through digital tools [15, 29], and supporting specific patient groups like children in healthcare settings [39]. These studies emphasize the need for context-sensitive, rapport-enhanced digital healthcare communication [7].

Advancements in human-AI collaboration have led to the use of NLP methods for enhancing communication. Techniques such as reinforcement learning [40] and LLMs [41, 49, 52] are increasingly being utilized as tools for empathetic message generation in medical fields, and assistance such as clinical record processing [2], symptom pre-screening [23] and clinical trial cohort selection [28]. Additionally, research in domains like online communication [12] and medicine [32] are exploring the integration of these AI-driven methodologies. Built upon these developments and insights, this work introduces a specialized LLM-based method, solely focused on enhancing rapport in mHealth communications with no compromization of clinical information delivered.

Table 1: Distribution of Communication Types by Sender and Recipient Roles. Note, the 'Rapport' column indicates the function of ONLY rapport-building with no information; the 'Information' column indicates the function of ONLY information-delivery with no rapport-building; the 'Both' column indicates both information-delivery and rapport-building are in the message.

Sender Distribution

Sender Role	Rapport	Info	Both	All
Patient	154	364	151	671
Provider	44	136	128	308
Staff	34	622	97	757

Recipient Distribution

Recipient	Rapport	Info	Both	All
Patient	78	750	225	1057
Provider	63	203	101	367
Staff	91	169	50	312

3 DATASET DESCRIPTION

Our study utilizes a dataset of 1,740 HIV care conversation messages collected through an mHealth platform, PositiveLinks[1], spanning from October 30, 2017, to May 29, 2018. The data collection study was reviewed and approved by the university's Institutional Review Board (IRB), and all patients provided written informed consent to participate. All messages were deidentified to protect patient privacy. The dataset was manually annotated by human coders to categorize messages topics (app-related, medical, or social concerns) and functions (information exchange or rapport-building).Specifically, 'information-only' messages included refill requests or appointment scheduling conversations focused on practical needs; 'rapport-building' messages included a psychosocial component, expression of emotion, or other strengthening of the relationship between sender and recipient. Annotations were performed using a standardized codebook with established inter-rater reliability, consistent with coding methods used in other studies of patient-clinician communication in secure messaging [24, 37, 44]. Each message was coded as a single unit of expression, which could have more than one code applied to it, if relevant. Utterances within each message were not split into smaller subunits but considered as a whole message. Some messages did not elicit any reply from their recipient, but others occurred in a back-and-forth conversation, depending on the question or topic being discussed.

The messages were sent and received by patients, providers, and PositiveLinks program staff. Patients include people with HIV receiving care at a Ryan White Clinic who are enrolled in PositiveLinks. Providers include HIV physicians, nurses, mental health clinicians, case managers, and other roles involved in HIV care for enrolled patients. Staff roles include customer support for the platform and managing the technical needs of the program. The distribution of messages based on sender and recipient roles across patients, providers, and staff is summarized in the side-by-side subtables of Table 1.

Regarding the **sender** role, *patients* show a considerable involvement in sending messages, including 154 rapport-building messages, 364 information-exchange messages, and a notable 151 messages serving both functions. This highlights their active engagement in both aspects of communication. *Providers*, sending 308 messages in total, demonstrate a significant overlap with 128 messages serving both rapport building and information exchange functions, reflecting their multifaceted role in patient interactions. *Program staff* predominantly engage in information exchange (622 messages), which is more than double their involvement in rapport building (34 messages), indicating their central role in disseminating information.

Conversely, in the role of **recipients**, *patients* are the primary recipients of messages, receiving 1,057 messages in total, with a substantial portion (225 messages) combining both information exchange and rapport-building elements. This underscores their active role in the communication process and the importance of addressing their informational and emotional needs. *Providers* and *program staff* also engage significantly as recipients, with program staff receiving 312 messages (50 combining both functions) and providers receiving 367 messages (101 combining both functions).

4 STUDY 1: LINGUISTIC PERSPECTIVES OF HIV MOBILE COMMUNICATION

This section delves into the linguistic intricacies of HIV care conversations between 'rapport-building' and 'information-only' messages. Understanding these nuances is pivotal for enhancing quality communication, forming the groundwork for the AI-driven analysis and assistance.

4.1 LIWC-Based Linguistic Featurization

We utilized the Linguistic Inquiry and Word Count (LIWC) tool to process 1,740 messages from the PositiveLinks platform, extracting 118 features from 'Word Count' to 'Emotional Tone.' These features are crucial for understanding the communicative dynamics in HIV care. A comprehensive list of LIWC features is available in [48].

4.2 Statistical Analysis

We conducted an independent two-sample t-test to compare LIWC features between rapport-building and information-only messages across four sender groups: all samples, and specifically patients, providers, and staff as senders[2]. Significant differences in linguistic patterns emerged, varying by group.

The analysis revealed 69 significant features in the overall sample, with distinct variances across sender groups: 66 in patients, 31 in providers, and 87 in staff, respectively. Specifically, as shown in Table 2, emotion-related features, like emotional tone and affective processes, were consistently among the most significant features across all groups, emphasizing the importance of emotional expression in healthcare communication. In particular, the analysis of patients as senders revealed a strong emphasis on social and prosocial language, reflecting their focus on building connections and support in their communications. Conversely, providers and staff showed a more varied set of significant features, indicating their

[1]PositiveLinks, https://www.positivelinks4ric.com

[2]Prior to executing this test, the Shapiro-Wilk and Levene's tests were used to validate the data's normality and equality of variances, respectively.

Table 2: Top 10 Significant LIWC Features of 'Rapport-Buidling' Messages Compared to 'Information-Only' Ones by Role Group with Directional Sign.

	Patients as Senders		Providers as Senders		Staff as Senders	
Rank	Feature	p-value	Feature	p-value	Feature	p-value
1	Emotional Tone (+)	$8.54e^{-45***}$	Emotional Tone (+)	$3.76e^{-21***}$	Affective Processes (+)	$1.98e^{-36***}$
2	Positive Tone (+)	$1.49e^{-35***}$	Affective Processes (+)	$4.43e^{-9***}$	Emotional Tone (+)	$5.80e^{-30***}$
3	Affective Processes (+)	$2.10e^{-34***}$	Positive Tone (+)	$4.48e^{-9***}$	Positive Tone (+)	$6.56e^{-30***}$
4	Social Processes (+)	$8.87e^{-29***}$	Exclamation (+)	$2.26e^{-6***}$	Allure (+)	$1.19e^{-27***}$
5	Prosocial Behavior (+)	$2.33e^{-27***}$	Emotion Expression (+)	$5.41e^{-6***}$	Linguistic (+)	$1.37e^{-25***}$
6	Politeness (+)	$1.21e^{-25***}$	Positive Emotion (+)	$6.16e^{-6***}$	Common Verbs (+)	$2.83e^{-22***}$
7	Clout (+)	$5.23e^{-23***}$	Words per Sentence (-)	$2.29e^{-4***}$	Exclamation (+)	$4.56e^{-21***}$
8	Social Referents (+)	$2.57e^{-20***}$	Technology Words (-)	$3.65e^{-4***}$	Function Words (+)	$8.60e^{-20***}$
9	Communication Words (+)	$7.17e^{-20***}$	Culture-Related (-)	$4.39e^{-4***}$	Emotion Expression (+)	$8.60e^{-20***}$
10	2nd Person (+)	$3.78e^{-17***}$	Determiners (-)	$8.96e^{-4***}$	Technology Words (-)	$3.20e^{-19***}$

multifaceted roles in communication, particularly including more use of exclamations, shorter sentences, and less technology-related language by providers, balancing between information delivery and emotional support.

5 STUDY 2: RAPPORT-BUILDING CLASSIFICATION

Building on Study 1's insights, Study 2 focuses on developing AI-driven classification models for binary identification of 'rapport-building' and 'information-only.' In this study, based on the messages sent by providers, we explored two approaches: 1) widely-used ML methods and 2) emerging LLMs[3].

5.1 Model and Prompt Settings

Regarding the **ML** approach, we applied supervised ML models to classify messages based on text data processed through TF-IDF vectorization (a common way to convert text into digital representation) and linguistic features derived from LIWC. To benchmark the models' performance across different data types (i.e., text data and linguistic features), we employed a range of ML models, including classical algorithms (i.e., Random Forest) and deep learning models (i.e., Multi-Layer Perceptron, Long Short-Term Memory Networks (LSTM); note, the LSTM is only compatible for TF-IDF data). We utilized 5-fold cross-validation for a robust evaluation, using Balanced Accuracy and Macro-Weighted F1 Score as performance metrics, each with a baseline of 50% for balanced assessment in imbalanced datasets.

Transitioning to the use of the **LLM**-based classification, we self-hosted and utilized the state-of-the-art, open-source Llama-2 model[4] (70B version). The linguistic understanding and reasoning capabilities of LLMs make them highly capable for text classification, even without domain-specific training. Our approach employs this model to classify messages using its pre-trained knowledge base without additional training. Our goal is to leverage a general-purpose LLM supplemented with linguistic hints from Study 1, to detect rapport building using the prompt structure:

$$Prompt_{identify} = MessageText + Prompt_{task} + Prompt_{linguistic\_insights} + OutputConstraint$$

Specifically, the initial setup involves a structured prompt comprising: 1) the provider's message text, 2) $Prompt_{task}$, a directive explaining the classification task, i.e., *"In this message from the provider between a HIV healthcare provider and a patient, is the content information-only or expressing rapport-building?"*, 3) initially excluding $Prompt_{linguistic\_insights}$, and 4) $OutputConstraint$ to define the LLM's response format, i.e., *"Answer in the format of either 'rapport-building' or 'information-only' "*. Subsequently, inspired by in-context engineering (aka. prompt engineering) [10, 53], we enhance the prompt with $Prompt_{linguistic_insights}$, i.e., *"Considering the linguistic cues in decision-making; focusing on elements like emotional tone, relational language, and the presence of supportive connection indicators."*, derived from Study 1's findings.

5.2 Classification Results

The results in Table 3 indicate compatible performance of zero-shot LLM approaches compared to the supervised ML models, with the added potential of adaptability since they were not trained with any new data (i.e., zero-shot). The supervised ML models, when using LIWC linguistic features, particularly excel, with the Random Forest model achieving the highest Balanced Accuracy and Macro-Weighted F1 Score. This underscores the effectiveness of incorporating linguistic analysis in traditional ML methods. The potential reason behind the performance gap between the deep learning models (e.g., particularly, LSTM) and random forest may be the small dataset, which is hard to drive the data-consuming deep learning models. Additionally, the LlaMa-2-70B model in a zero-shot setting demonstrates a notable performance boost when enhanced with linguistic prompts derived from Study 1. The increase in both Balanced Accuracy and F1 Score with linguistic prompts suggests that the incorporation of specific linguistic hints into the prompts significantly augments the LLM's capability to classify messages.

6 STUDY 3: LLM-ENHANCED RAPPORT BUILDING IN HIV MHEALTH

In Study 3, we advance towards developing an LLM-based system capable of 1) identifying 'information-only' messages from the

[3]Note, despite being de-identified, the dataset still includes some sensitive Protected Health Information (PHI); therefore, we only utilized self-hosted start-of-the-art LLMs such as LlaMa2 rather than closed-source models like GPT-3.5 or GPT-4, ensuring data security and privacy compliance.
[4]Introducing LlaMa 2: https://ai.meta.com/llama/

Table 3: Performance Comparison of Different Models. To ensure fair comparisons, 5-fold cross-validation was performed for both supervised ML and zero-shot LLM methods using identical sets of randomly selected samples. The performance metrics for each method category were averaged over the five folds, and standard deviations were calculated to assess variability.

Method Type	Model	Balanced Accuracy	Macro-Weighted F1 Score
Baseline	Random Guess	0.500 ± 0.000	0.500 ± 0.000
Supervised ML	LSTM (TEXT TF-IDF)	0.506 ± 0.032	0.509 ± 0.027
	MLP (TEXT TF-IDF)	0.747 ± 0.049	0.746 ± 0.050
	Random Forest (TEXT TF-IDF)	0.700 ± 0.056	0.691 ± 0.070
	MLP (Liguistic LIWC)	0.770 ± 0.079	0.769 ± 0.085
	Random Forest (Linguistic LIWC)	0.812 ± 0.067	0.810 ± 0.069
Zero-Shot LLM	LlaMa-2-70B	0.751 ± 0.053	0.752 ± 0.062
	Prompt-Enhanced LlaMa-2-70B	**0.838 ± 0.075**	**0.832 ± 0.068**

Table 4: Comparative Analysis of Linguistic and Information Delivery Metrics in Different Rewriting Scenarios. The percentages in the tables indicate the increase (+) or decrease (-) by the original messages (↑ indicates higher is better, ↓ indicates lower is better). For the LIWC linguistic features, '*' indicates statistical significance in the paired sample t-tests comparing between original and rewritten messages (* p<0.05, ** p<0.01, *** p<0.001). Abbreviations: Tone=Emotional Tone, Affective=Affective Processes, Social=Social Processes.

		LIWC Linguistic Enhancement			Information Integrity	
Setting	Prompt	Tone (↑)	Affective (↑)	Social (↑)	Specificity (↑)	Edit Rate (↓)
Semi-Auto	$Prompt_{S1}$	+192.95%***	+220.23%***	+147.74%*	- 12.84%	+106.13%
	$Prompt_{S1\_LE}$	+206.40%***	+318.26%***	+212.76%***	-14.40%	+113.82%
Autonomous	$Prompt_{S2}$	+112.75%***	+135.70%***	+88.20%	-5.99%	+51.90%
	$Prompt_{S2\_LE}$	+121.31%***	+185.49%***	+130.74%*	-6.80%	+63.69%

providers that could be transformed into rapport-building messages, and 2) rewriting these messages to be enhanced in building rapport, while preserving the original information's integrity.

6.1 Evaluation and Prompt Settings

The basic prompt $Prompt_{rewrite}$ of rewriting was designed as "*Rewrite the message to build better patient-provider rapport, ensuring minimal edits and maintaining key information's integrity and specificity.*"

To assess the feasibility of LLMs in this context, we explore two varying settings:

- **Setting 1 (Human-Intervened Rewriting):** LLMs are tasked with rewriting messages labelled as 'information-only' by humans. This tests the LLM's general performance in enhancing rapport in provider-patient communication. To achieve this, the $Prompt_{S1}$ is applied to all provider messages labeled as 'information-only':

$$Prompt_{S1} = MessageText + Prompt_{rewrite}$$

- **Setting 2 (Autonomous Operation):** Eliminating human intervention, LLMs are tasked with autonomously identifying and rewriting 'information-only' messages into rapport-building ones. This simulates a real-world scenario where the LLM-based system operates independently, showcasing its potential for automated rapport enhancement in clinical communication. The $Prompt_{S2}$ is also applied to all 'information-only' messages:

$$Prompt_{S2} = MessageText + Prompt_{identify} + \text{'if information-only,'} + Prompt_{rewrite}$$

For both $Prompt_{S1}$ and $Prompt_{S2}$, we also incorporated linguistic enhancements into the prompts based on insights from Study 1, i.e., *"Focus on enhancing elements like emotional tone, positive expressions, social language, and expressions that demonstrate empathy and understanding"*. The modified versions, $Prompt_{S1_LE}$ and $Prompt_{S2_LE}$, are thus designed to be more linguistically specific and enhanced.

For the evaluation metrics, we design them from the perspectives of 1) change in typical LIWC features such as ***emotional tone***, ***affective processes***, and ***social processes*** and 2) information integrity and delivery measured by ***information specificity*** and ***edit rate***, inspired by [40]. In specific:

- **Information Specificity**, computed using BERT embedding (high-dimensional vector representations of text) [13], assesses how closely a rewritten message aligns with the original content based on cosine similarity.
- **Edit Rate** measures the extent of text modification by calculating the Levenshtein distance ratio to the original text's length [31], indicating the balance between content's integrity and rapport-building modifications.

6.2 Rewriting Results

The results presented in Table 4 reflect the effectiveness of LLM-based rewriting in HIV mHealth communication. The performance metrics show that LLMs, particularly when guided by linguistic-enhanced prompts, excel in enhancing the emotional, affective, and social aspects of the messages, tested statistically significant by paired t-test between original and rewritten messages. This indicates a successful enhancement of rapport-building elements. Additionally, there is a natural trade-off between enhancing the linguistic aspects of messages and maintaining information integrity and delivery, as measured by information specificity and edit rate.

Among the options, the autonomous, linguistic-guided setting strikes a balance, improving rapport-building with minimal intervention, indicated by the significant rises in LIWC features, including emotional tone (121.31%), affective processes (185.49%), and social processes (130.74%), alongside a manageable loss in information specificity (6.80%) and an edit rate of 63.69%, compared to the original text. This highlights the LLM's potential in effectively identifying and enhancing 'information-only' messages, enhancing rapport while preserving the core information.

7 DISCUSSION

7.1 Implications

This pilot, preliminary study demonstrates the significant potential of integrating linguistic analysis with LLMs in enhancing the quality of mHealth communications in HIV care. The ability of these models to classify and rewrite messages to bolster rapport-building suggests a scalable paradigm in patient-provider interactions, where AI augments human communication for greater emotional and psychological support. These findings also resonate with existing literature that highlights the efficacy of LLMs in handling nuanced, subjective tasks involving identification and rewriting [27, 42, 43].

Practically, this research offers a novel approach for healthcare providers to improve their communication with patients in mHealth platforms. Implementing LLMs in mHealth apps can lead to more engaging and supportive conversations, potentially improving patient outcomes. This approach is particularly vital in HIV care and, more broadly, palliative care [9] and other contexts of serious illness, where effective communication can significantly impact treatment adherence and patient well-being.

Studies of patient-clinician communication demonstrate that both informational and emotional content of interactions are important [47]. Chronic disease management can be enhanced by the exchange of practical information between patients and clinicians through secure messaging [37]. AI-driven methods that favor rapport at the expense of information could reduce clarity of messages and cause misunderstandings. Instead, the goal of message rewriting to include more rapport is intended to reduce missed opportunities to meet patient needs. Studies have shown gaps in clinician responses to patient emotional cues [26], which AI assistance could help address. Our method's AI-enhanced messages have not yet been evaluated by HIV clinicians and patients using the messaging, which will be a key step needed in future work.

This study bridges traditional mobile health communication models with AI-driven methods, suggesting a promising shift towards integrating AI in healthcare dialogues. The enhanced performance seen with linguistically informed prompts underscores the value of combining human linguistic expertise or guidelines with language models.

7.2 Ethical and Privacy Issues

This study follows ethical, privacy, and data security protocols, ensuring robust protection of patient and provider information. Protected health information (PHI) is de-identified and not shared with third parties.

We advocate for healthcare policies to adopt such technologies, ensuring ethical use, data privacy, and effective AI tool training for healthcare professionals. This should include opt-in-and-out standards and compliance with guidelines like HIPAA [1]. Specifically, participation and informed consent for both patients and providers should be voluntary, focusing on ensuring understanding and avoiding any form of coercion, particularly for marginalized groups. Data security is crucial, with stringent encryption and anonymization to comply with regulations. Lastly, we stress the importance of analytical accuracy in AI outputs to prevent misinterpretation and uphold patient care integrity.

7.3 Limitations and Future Work

As a preliminary quantitative study, this work has several limitations which highlight future research opportunities. The pilot dataset's relatively small scale necessitates larger datasets in future studies. An immediate next step is developing an interactive user interface (UI) for interventions and a real-time suggestion pop-up system. Involving stakeholders like patients and clinicians for qualitative feedback will improve rewriting strategies and UI adaptability. Additionally, this preliminary, proof-of-concept work only tested the feasibility of zero-shot LLMs with no training based on labeled data, leaving space for supervised fine-tuning of LLMs to boost model performance in both detection and rewriting. The quality of patient-provider relationships plays an important role in care engagement and clinical outcomes. Future research will include investigation of how AI-enhanced communication can assist clinicians in providing rapport in data-driven ways without undermining their unique connections with their patients.

8 CONCLUSION

This paper introduces an LLM-based approach, informed by linguistic analysis (study 1), to identify (study 2) and rewrite (study 3) 'information-only' messages from HIV mHealth providers into 'rapport-building' communications. Using a pilot dataset from HIV mHealth interactions demonstrates the successful application of LLMs in healthcare communication, effectively balancing linguistic enhancement with information integrity. This approach highlights the feasibility of AI to improve the quality and effectiveness of digital healthcare interactions.

REFERENCES

[1] Accountability Act. 1996. Health insurance portability and accountability act of 1996. *Public law* 104 (1996), 191.
[2] Monica Agrawal, Stefan Hegselmann, Hunter Lang, Yoon Kim, and David Sontag. 2022. Large language models are few-shot clinical information extractors. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*. 1998–2022.

[3] Jordan M Alpert, Chelsea N Hampton, Merry Jennifer Markham, and Carma L Bylund. 2022. Clinicians' attitudes and behaviors towards communicating electronically with patients: a grounded practical theory approach. *Journal of Health Communication* 27, 2 (2022), 103–114.
[4] Ashley Archiopoli, Tamar Ginossar, Bryan Wilcox, Magdalena Avila, Ricky Hill, and John Oetzel. 2016. Factors of interpersonal communication and behavioral health on medication self-efficacy and medication adherence. *AIDS care* 28, 12 (2016), 1607–1614.
[5] Mary Catherine Beach, Jeanne Keruly, and Richard D Moore. 2006. Is the quality of the patient-provider relationship associated with better adherence and health outcomes for patients with HIV? *Journal of general internal medicine* 21 (2006), 661–665.
[6] Mary Catherine Beach, Debra L Roter, Somnath Saha, P Todd Korthuis, Susan Eggly, Jonathan Cohn, Victoria Sharp, Richard D Moore, and Ira B Wilson. 2015. Impact of a brief patient and provider intervention to improve the quality of communication about medication adherence among HIV patients. *Patient education and counseling* 98, 9 (2015), 1078–1083.
[7] Robert Bowman, Camille Nadal, Kellie Morrissey, Anja Thieme, and Gavin Doherty. 2023. Using Thematic Analysis in Healthcare HCI at CHI: A Scoping Review. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–18.
[8] Meredith L Brantley, Deanna Kerrigan, Danielle German, Sahnah Lim, and Susan G Sherman. 2017. Identifying patterns of social and economic hardship among structurally vulnerable women: a latent class analysis of HIV/STI risk. *AIDS and Behavior* 21 (2017), 3047–3056.
[9] Lisa Jane Brighton and Katherine Bristowe. 2016. Communication in palliative care: talking about the end of life, before the end of life. *Postgraduate medical journal* 92, 1090 (2016), 466–470.
[10] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems* 33 (2020), 1877–1901.
[11] Chelsea E Canan, Marika E Waselewski, Ava Lena D Waldman, George Reynolds, Tabor E Flickinger, Wendy F Cohn, Karen Ingersoll, and Rebecca Dillingham. 2020. Long term impact of PositiveLinks: Clinic-deployed mobile technology to improve engagement with HIV care. *PLoS One* 15, 1 (2020), e0226870.
[12] Jan Clusmann, Fiona R Kolbinger, Hannah Sophie Muti, Zunamys I Carrero, Jan-Niklas Eckardt, Narmin Ghaffari Laleh, Chiara Maria Lavinia Löffler, Sophie-Caroline Schwarzkopf, Michaela Unger, Gregory P Veldhuizen, et al. 2023. The future landscape of large language models in medicine. *Communications Medicine* 3, 1 (2023), 141.
[13] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).
[14] Rebecca Dillingham, Karen Ingersoll, Tabor E Flickinger, Ava Lena Waldman, Marika Grabowski, Colleen Laurence, Erin Wispelwey, George Reynolds, Mark Conaway, and Wendy F Cohn. 2018. PositiveLinks: a mobile health intervention for retention in HIV care and clinical outcomes with 12-month follow-up. *AIDS patient care and STDs* 32, 6 (2018), 241–250.
[15] Xianghua Ding, Yunan Chen, Zhaofei Ding, and Yiwen Xu. 2019. Boundary negotiation for patient-provider communication via wechat in china. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–24.
[16] Xianghua Ding, Xinning Gui, Xiaojuan Ma, Zhaofei Ding, and Yunan Chen. 2020. Getting the Healthcare We Want: The Use of Online" Ask the Doctor" Platforms in Practice. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.
[17] Michael H Farrell and Stephanie A Christopher. 2013. Frequency of high-quality communication behaviors used by primary care providers of heterozygous infants after newborn screening. *Patient education and counseling* 90, 2 (2013), 226–232.
[18] Tabor E Flickinger, Karen Ingersoll, Sabrina Swoger, Marika Grabowski, and Rebecca Dillingham. 2020. Secure messaging through PositiveLinks: examination of electronic communication in a clinic-affiliated smartphone app for patients living with HIV. *Telemedicine and e-Health* 26, 3 (2020), 359–364.
[19] Tabor E Flickinger, Somnath Saha, Richard D Moore, and Mary C Beach. 2013. Higher quality communication and relationships are associated with improved patient engagement in HIV care. *JAIDS Journal of Acquired Immune Deficiency Syndromes* 63, 3 (2013), 362–366.
[20] Tabor E Flickinger, Somnath Saha, Debra Roter, P Todd Korthuis, Victoria Sharp, Jonathan Cohn, Susan Eggly, Richard D Moore, and Mary Catherine Beach. 2016. Clinician empathy is associated with differences in patient–clinician communication behaviors and higher medication self-efficacy in HIV care. *Patient education and counseling* 99, 2 (2016), 220–226.
[21] Craig H Ganoe, Weiyi Wu, Paul J Barr, William Haslett, Michelle D Dannenberg, Kyra L Bonasia, James C Finora, Jesse A Schoonmaker, Wambui M Onsando, James Ryan, et al. 2021. Natural language processing for automated annotation of medication mentions in primary care visit conversations. *JAMIA open* 4, 3 (2021), ooab071.
[22] James L Graham, Lokesh Shahani, Richard M Grimes, Christine Hartman, and Thomas P Giordano. 2015. The influence of trust in physicians and trust in the healthcare system on linkage, retention, and adherence to HIV care. *AIDS patient care and STDs* 29, 12 (2015), 661–667.
[23] Danny M den Hamer, Perry Schoor, Tobias B Polak, and Daniel Kapitan. 2023. Improving Patient Pre-screening for Clinical Trials: Assisting Physicians with Large Language Models. *arXiv preprint arXiv:2304.07396* (2023).
[24] Timothy P Hogan, Tana M Luger, Julie E Volkman, Mary Rocheleau, Nora Mueller, Anna M Barker, Kim M Nazi, Thomas K Houston, and Barbara G Bokhour. 2018. Patient centeredness in electronic communication: evaluation of patient-to-health care team secure messaging. *Journal of medical Internet research* 20, 3 (2018), e82.
[25] Jeremy Howick, Andrew Moscrop, Alexander Mebius, Thomas R Fanshawe, George Lewith, Felicity L Bishop, Patriek Mistiaen, Nia W Roberts, Eglė Dieninytė, Xiao-Yang Hu, et al. 2018. Effects of empathic and positive communication in healthcare consultations: a systematic review and meta-analysis. *Journal of the Royal Society of Medicine* 111, 7 (2018), 240–252.
[26] Ian Hsu, Somnath Saha, Phillip Todd Korthuis, Victoria Sharp, Jonathon Cohn, Richard D Moore, and Mary Catherine Beach. 2012. Providing support to patients in emotional encounters: a new perspective on missed empathic opportunities. *Patient education and counseling* 88, 3 (2012), 436–442.
[27] Fan Huang, Haewoon Kwak, and Jisun An. 2023. Is chatgpt better than human annotators? potential and limitations of chatgpt in explaining implicit hate speech. *arXiv preprint arXiv:2302.07736* (2023).
[28] Qiao Jin, Zifeng Wang, Charalampos S Floudas, Jimeng Sun, and Zhiyong Lu. 2023. Matching patients to clinical trials with large language models. *ArXiv* (2023).
[29] Ha-Kyung Kong and Karrie Karahalios. 2020. Addressing cognitive and emotional barriers in parent-clinician communication through behavioral visualization webtools. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–12.
[30] Colleen Laurence, Erin Wispelwey, Tabor E Flickinger, Marika Grabowski, Ava Lena Waldman, Erin Plews-Ogan, Claire Debolt, George Reynolds, Wendy Cohn, Karen Ingersoll, et al. 2019. Development of PositiveLinks: a mobile phone app to promote linkage and retention in care for people with HIV. *JMIR formative research* 3, 1 (2019), e11578.
[31] Vladimir I Levenshtein et al. 1966. Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady*, Vol. 10. Soviet Union, 707–710.
[32] Siru Liu, Allison B McCoy, Aileen P Wright, Babatunde Carew, Julian Z Genkins, Sean S Huang, Josh F Peterson, Bryan Steitz, and Adam Wright. 2023. Leveraging large language models for generating responses to patient messages. *medRxiv* (2023), 2023–07.
[33] Stewart W Mercer, Melanie Neumann, Markus Wirtz, Bridie Fitzpatrick, and Gaby Vojt. 2008. General practitioner empathy, patient enablement, and patient-reported outcomes in primary care in an area of high socio-economic deprivation in Scotland—a pilot prospective study using structural equation modeling. *Patient education and counseling* 73, 2 (2008), 240–245.
[34] Diane S Morse, Elizabeth A Edwardsen, and Howard S Gordon. 2008. Missed opportunities for interval empathy in lung cancer communication. *Archives of internal medicine* 168, 17 (2008), 1853–1858.
[35] Jean B Nachega, Chelsea Morroni, José M Zuniga, Mauro Schechter, Jürgen Rockstroh, Suniti Solomon, and Renslow Sherer. 2012. HIV treatment adherence, patient health literacy, and health care provider–patient communication: Results from the 2010 AIDS Treatment for Life International Survey. *Journal of the International Association of Physicians in AIDS care* 11, 2 (2012), 128–133.
[36] Jenny Park, Somnath Saha, Dingfen Han, Stéphanie De Maesschalck, Richard Moore, Todd Korthuis, Debra Roter, Amy Knowlton, Tanita Woodson, and Mary Catherine Beach. 2019. Emotional communication in HIV care: an observational study of patients' expressed emotions and clinician response. *AIDS and Behavior* 23 (2019), 2816–2828.
[37] Stephanie A Robinson, Mark Zocchi, Carolyn Purington, Linda Am, Kathryn DeLaughter, Varsha G Vimalananda, Dane Netherton, Arlene S Ash, Timothy P Hogan, and Stephanie L Shimada. 2023. Secure Messaging for Diabetes Management: Content Analysis. *JMIR diabetes* 8 (2023), e40272.
[38] Padhraig Ryan, Saturnino Luz, Pierre Albert, Carl Vogel, Charles Normand, and Glyn Elwyn. 2019. Using artificial intelligence to assess clinicians' communication skills. *Bmj* 364 (2019).
[39] Woosuk Seo, Ayse G Buyuktur, Sanya Verma, Hyeryoung Kim, Sung Won Choi, Laura Sedig, and Sun Young Park. 2021. Learning from Healthcare Providers' Strategies: Designing Technology to Support Effective Child Patient-Provider Communication. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–15.
[40] Ashish Sharma, Inna W Lin, Adam S Miner, David C Atkins, and Tim Althoff. 2021. Towards facilitating empathic conversations in online mental health support: A reinforcement learning approach. In *Proceedings of the Web Conference 2021*. 194–205.
[41] Ashish Sharma, Inna W Lin, Adam S Miner, David C Atkins, and Tim Althoff. 2023. Human–AI collaboration enables more empathic conversations in text-based peer-to-peer mental health support. *Nature Machine Intelligence* 5, 1 (2023),

46–57.

[42] Ashish Sharma, Kevin Rushton, Inna Wanyin Lin, Theresa Nguyen, and Tim Althoff. 2023. Facilitating Self-Guided Mental Health Interventions Through Human-Language Model Interaction: A Case Study of Cognitive Restructuring. *arXiv preprint arXiv:2310.15461* (2023).

[43] Ashish Sharma, Kevin Rushton, Inna Wanyin Lin, David Wadden, Khendra G Lucas, Adam S Miner, Theresa Nguyen, and Tim Althoff. 2023. Cognitive Reframing of Negative Thoughts through Human-Language Model Interaction. *arXiv preprint arXiv:2305.02466* (2023).

[44] Stephanie L Shimada, Beth Ann Petrakis, James A Rothendler, Maryan Zirkle, Shibei Zhao, Hua Feng, Gemmae M Fix, Mustafa Ozkaynak, Tracey Martin, Sharon A Johnson, et al. 2017. An analysis of patient-provider secure messaging at two Veterans Health Administration medical centers: message content and resolution through secure messaging. *Journal of the American Medical Informatics Association* 24, 5 (2017), 942–949.

[45] Simone J Skeen, Stephen Scott Jones, Carolyn Marie Cruse, and Keith J Horvath. 2022. Integrating natural language processing and interpretive thematic analyses to gain human-centered design insights on HIV mobile health: Proof-of-concept analysis. *JMIR Human Factors* 9, 3 (2022), e37350.

[46] Irena Spasic, Goran Nenadic, et al. 2020. Clinical text data in machine learning: systematic review. *JMIR medical informatics* 8, 3 (2020), e17984.

[47] Richard L Street Jr. 2013. How clinician–patient communication contributes to health improvement: modeling pathways from talk to outcome. *Patient education and counseling* 92, 3 (2013), 286–291.

[48] Yla R Tausczik and James W Pennebaker. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology* 29, 1 (2010), 24–54.

[49] Arun James Thirunavukarasu, Darren Shu Jeng Ting, Kabilan Elangovan, Laura Gutierrez, Ting Fang Tan, and Daniel Shu Wei Ting. 2023. Large language models in medicine. *Nature medicine* 29, 8 (2023), 1930–1940.

[50] James A Tulsky, Mary Catherine Beach, Phyllis N Butow, Susan E Hickman, Jennifer W Mack, R Sean Morrison, Richard L Street, Rebecca L Sudore, Douglas B White, and Kathryn I Pollak. 2017. A research agenda for communication between health care professionals and patients living with serious illness. *JAMA internal medicine* 177, 9 (2017), 1361–1366.

[51] Juddy Wachira, Ann Mwangi, Diana Chemutai, Monica Nyambura, Becky Genberg, and Ira B Wilson. 2021. Higher Clinician-Patient Communication Is Associated With Greater Satisfaction With HIV Care. *Journal of the International Association of Providers of AIDS Care (JIAPAC)* 20 (2021), 23259582211054935.

[52] Haoyi Xiong, Jiang Bian, Sijia Yang, Xiaofei Zhang, Linghe Kong, and Daqing Zhang. 2023. Natural Language based Context Modeling and Reasoning for Ubiquitous Computing with Large Language Models: A Tutorial. arXiv:2309.15074 [cs.CL]

[53] Yongchao Zhou, Andrei Ioan Muresanu, Ziwen Han, Keiran Paster, Silviu Pitis, Harris Chan, and Jimmy Ba. 2022. Large language models are human-level prompt engineers. *arXiv preprint arXiv:2211.01910* (2022).

[54] Kelly B Haskard Zolnierek and M Robin DiMatteo. 2009. Physician communication and patient adherence to treatment: a meta-analysis. *Medical care* 47, 8 (2009), 826.