

Understanding the Impact of Long-Term Memory on Self-Disclosure with Large Language Model-Driven Chatbots for Public Health Intervention

Jo, E., Jeong, Y., Park, S., Epstein, D. A., & Kim, Y. H. (2024). Understanding the impact of long-term memory on self-disclosure with large language model-driven chatbots for public health intervention. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (CHI '24), May 11–16, 2024, Honolulu, HI, USA. ACM.

INTRODUCTION

- 公衆衛生モニタリングは、感染症の拡大を抑え、地域社会の健康と福祉を守る上で重要な役割を果たす (48, 69)
 - 市民のウェルビーイングに関する日常的な習慣への意識向上においても重要である (20, 66)
- 公衆衛生モニタリングは、人々からの定期的なデータ収集が必要とされ、公衆衛生当局にとっては膨大な時間と労力を要することが多い (26–28, 53)

INTRODUCTION

- COVID-19の際には、ある報道記事で、接触追跡者が1件の感染事例ごとに約10人の接触者に電話をかけ、その後2日ごとに追跡を行っていた
 - 米国の50州のうち37州が、初期段階で感染拡大を効果的に抑えるための十分な追跡者を確保できていなかったと推定されている⁽⁶⁷⁾
- 大規模なモニタリングにおける公衆衛生当局の負担を軽減するために、チャットボットやモバイルアプリなどのテクノロジーがますます利用・提案されている
 - これには、接触追跡^(35, 47, 48)、母子健康教育^(26-28, 53, 79)、および社会的孤立の介入⁽³⁰⁾が含まれる

INTRODUCTION

- 課題:いかにして効果的に健康情報を開示させるか
 - 人々は自分の弱みをさらけ出すことを避ける傾向があり(13, 73)、そのため健康問題などのデリケートな情報を開示することにためらいを感じることが多い
 - 公衆衛生モニタリングでは、定期的な関与を維持することが求められるが、長期間にわたって健康情報の開示を促進し続けることは難しい(39, 44)

INTRODUCTION

- チャットボットによる公衆衛生モニタリング
 - メリット
 - チャットボットはデリケートな話題について自己開示を引き出すのに効果的である可能性が示されている(18, 43, 50, 60)
 - 課題
 - 現在のLLM駆動のチャットボットは、過去のセッションから得た情報を記憶し参照する機能がほとんどないため、過去の会話に基づいた個人の健康履歴をフォローアップするのに苦労することが多く、それがユーザーとの関与や、モニタリングに必要な自己開示の減少につながる可能性がある(30)

INTRODUCTION

- 自然言語処理(NLP)の研究では、大規模言語モデル(LLM)に長期記憶(LTM)を付加する試みが進んでいる
 - LTMは、個人の興味や健康状態などの情報を、現在の会話セッションを超えて記憶する能力を指す
 - LTMを持つLLM駆動のチャットボットは、各セッションから特定のトピックに関する情報を保存し、次回の会話で他の入力データと一緒にモデルに提供することで、チャットボットが参照するための手がかりを提供する

INTRODUCTION

- OpenAIのChatGPT⁽⁵⁴⁾やGoogleのBard⁽¹⁹⁾などの「メモリ」・「カスタム指示」機能
 - これらの機能は、特定のユーザーの背景情報やスタイルの好みをセッションを超えて静的に保持し、チャットボットの挙動に反映させる⁽⁵⁵⁾
- 公衆衛生の文脈では、ユーザーの健康状態は絶えず変化するため、メモリを動的に更新することが重要

INTRODUCTION

- LTMを持つLLM駆動のチャットボット
 - 過去のセッションから得た情報を動的に保存、更新、参照するように設計されることで、個人の具体的な健康問題に焦点を当てたパーソナライズされた会話を提供する
 - 自己開示を促進する可能性がある
 - 公衆衛生の分野でユーザーとの関与を維持するために有益であり、これは地域社会の健康と福祉を理解する上で不可欠である
 - LTMが公衆衛生介入においてユーザーとの対話やチャットボットの印象にどのような影響を与えるかについての理解は限られている

INTRODUCTION

- RQ: LTMは公衆衛生介入の文脈において、ユーザーの自己開示とLLM駆動のチャットボットに対する印象にどのような影響を与えるのか？

CareCall

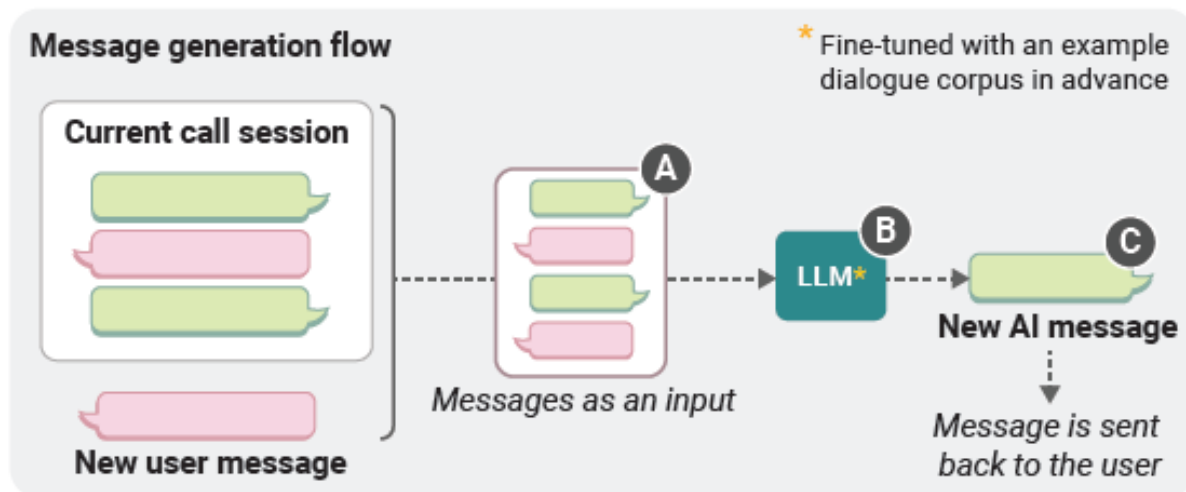
- CareCallによる公衆衛生モニタリング
 - 社会的に孤立した個人の健康状態を確認する電話を通じてモニタリングを行う、LLM駆動の音声チャットボット
 - 定期的に(週に1, 2回など)やり取りを行った
 - 公衆衛生モニタリングを支援するために設計され、韓国の地方自治体の公衆衛生機関と連携して使用された
 - 異常シグナルを検知した場合、職員が個別に連絡をした
 - 当初、CareCallはLTMを備えていなかったが、後にこの機能が統合された
 - LTMなし期間: 2021年11月から2022年9月

CareCall

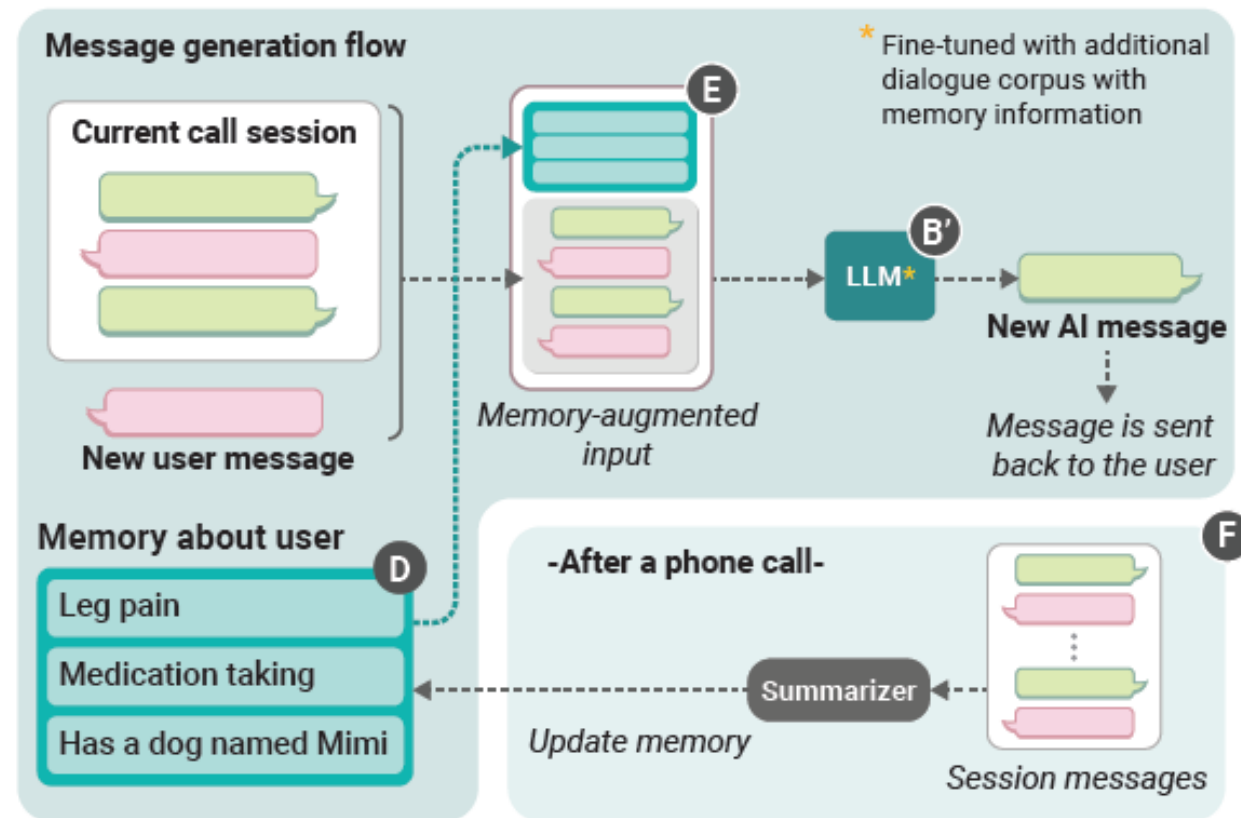
- CareCallとLTM

- 各セッション終了後、5つのLTMトピックに関連する要約文を生成しメモリへ記録
 - 健康、食事、睡眠、訪れた場所、ペットを以下の優先順位でキューレーション
 - 1. 健康 2. 食事と睡眠 3. 訪れた場所とペット
- CareCallは次回以降のセッションでLTMをトリガーとする質問を提示
 - 例) ユーザーが「足の痛みで医師にかかっている」と話した場合、後のセッションで「足の調子はどうですか？」と質問
 - しかし、後に「治った」と報告されれば、その情報は削除
- LTMイベントが発生するかどうかは会話の内容に依存
 - 必ずしもLTMイベントが発生するとは限らない

CareCall



(a) CareCall without LTM



(b) CareCall with LTM

METHODOLOGY

- コールログ
 - 混合研究法 (mixed-method approach) を使用し、量的および質的にログを解析
- インタビュー
 - LTMを使用した9人のユーザーにインタビューを実施

METHODOLOGY

- CareCallの実際の運用から得た1,252件のコールログ
 - 10週間分のログを収集
 - 中・高年層の独居生活者(40代～)
 - LTMあり群: LTMを使用するCareCallを最初から使用したユーザー
 - LTMなし群: LTMなしのCareCallのみを使用したユーザー

METHODOLOGY

- スクリーニング

- LTMあり群の内、LTMイベントが最低1回起きたユーザー
- 各グループで4週間内に2回以上電話に応答したユーザー

- 最終的なデータセット

- LTMあり群: 66名 (A市35名、B市31名)、男性17名、女性49名
 - 576件のコールログ
- LTMなし群: 81名 (A市39名、B市42名)、男性32名、女性49名
 - 676件のコールログ

METHODOLOGY

- LTMを10か月以上、週1回使用したユーザーを対象にインタビューを実施
 - インタビューは電話で行い、1回あたり20～30分で録音
 1. CareCallとの印象的な会話（ポジティブまたはネガティブ）
 2. LTMイベントの体験（エージェントが記憶した情報やそのフォローアップの印象）
 3. 必要な情報を忘れたケース
 4. エージェントに記憶してほしい情報や記憶してほしくない情報

Table 1: Demographics of interview participants and duration of CareCall use at the point of the interviews

Alias	Age	Gender	Duration of CareCall Use
P1	65	Male	10 months
P2	66	Male	10 months
P3	61	Male	10 months
P4	61	Female	10 months
P5	65	Male	10 months
P6	76	Male	4 months
P7	66	Female	10 months
P8	77	Male	10 months
P9	81	Female	10 months

Data Analysis Quantitative Analysis: Call Logs

- コールログの定量分析
 - コールログをAIによって書き起こした
 - 書き起こされたコールログからコードブックを作成した
 - コード数を従属変数とした

Data Analysis Quantitative Analysis: Call Logs

- コードブックの作成
 - まず、筆頭著者が40名のユーザーのコールログをオープンコーディング
 - その後、研究チーム全体で定期的に議論しながら最終的なコードブックを作成
 - 最終的に9つのカテゴリと19のコードから構成

Data Analysis

B FINAL CODEBOOK FOR QUALITATIVE ANALYSIS OF CALL LOGS

Category	Code	Code description	Example
Meals	<i>Meals-simple</i>	Whether a person has eaten, or whether they have an appetite	"Yeah, I just had lunch." "I don't really feel like eating."
	<i>Meals-detail</i>	What kind of food a person had, or why they have difficulty eating	"I had Kimchi Stew for lunch." "I've not been feeling well, so I lost my appetite."
Sleep	<i>Sleep-simple</i>	Whether a person sleeps well	"Yeah, I'm having a good sleep."
	<i>Sleep-detail</i>	Details on sleep quality, including why a person has difficulty sleeping	"I have a hard time falling asleep." "I often wake up in the middle of the night." "I couldn't sleep well because of the back pain."
Health	<i>Health-simple</i>	Whether a person is experiencing any health issues that likely warrant clinical care/attention	"I was a bit under the weather yesterday, but now I'm doing better." "I'm feeling well."
	<i>Health-detail</i>	Details about one's health issues, including physical/mental discomforts, pain, or clinical symptoms.	"I have a toothache." "I have leg pain, so I can't walk around much." "I have been terribly ill over the last few days." "My diabetes got worse"
	<i>Health-attitude</i>	Thoughts, feelings, and attitudes toward their health and clinical treatment	"I make regular visits to my doctor but it isn't really working." "I think I'll just have to live with my leg pain." "I'm too old to get better."
Clinical	<i>Clinical-simple</i>	Whether a person is currently seeking clinical care or taking medications regularly, and whether they recently saw a doctor	"Yeah, I've been seeing a doctor." "(Are you taking your pills regularly?) Yes."
	<i>Clinical-detail</i>	Details about clinical care being sought (e.g., treatment, medical specialists)	"I'm taking pills for high blood pressure." "I have a slipped disc." "I just went to see a dentist." "I got pain relief shots on my back."
Activity	<i>Activity-simple</i>	Whether a person engages in any social or physical activities	"I'm just staying home today." "I'm just chilling at home." "I'm just watching TV."
	<i>Social-detail</i>	Details about one's social activities, including people, place, and jobs	"I just came back from the community center." "I'm at my daughter's place." "I'm working as a babysitter."
	<i>Physical-detail</i>	Details about one's physical activities, including chores and exercises, reasons if not engaged in any	"I'm doing some laundry now." "I'm taking some aerobic classes." "I can't do any exercise 'cause I have back pain." "I'm not doing any exercise. Just in my wheelchair all day."
Wellbeing	<i>Wellbeing-simple</i>	How a person is doing in general	"I'm doing okay." "I'm a bit tired today."
	<i>Wellbeing-detail</i>	Details about one's general wellbeing in both physical and mental aspects	"I usually feel a bit tired in the morning." "I think I need some rest today." "Life is not enjoyable living alone. Don't have anyone around."
Hobbies	<i>Hobbies</i>	Descriptions of one's interests, preferences, and hobbies	"I like to drink coffee with sugar." "I like to read detective fiction." "I'm transcribing the Bible."
Positive Reactions	<i>Appreciation</i>	A user thanks the agent explicitly. Conventional expressions of thanks in farewell are not included.	"It was really nice of you to say that." "Thanks for looking out for me." "Thank you for checking in with me."
	<i>Anthromorphization</i>	A user interacts with the agent as if it were a human being.	"I'm doing well. How are you doing?" "By the way, what's your name?" "You have a great day." "I'll cook noodles for you one day."
Negative Reactions	<i>Negative Feedback</i>	A user expresses frustration to the agent.	"You're not a human. You're what, an answering machine?" "Stop playing the recording." "It's definitely not a human voice." "You're just playing a recorded message, right?"
	<i>Disregard</i>	A user does not respond to the agent's question, or gives yes/no answers to open questions, or hangs up the call without explanation.	"(What did you have for lunch?) Yeah." "(Do you have any plans for today?)" [hung up]

Data Analysis Quantitative Analysis

- 線形混合効果モデル

- 群間での自己開示の深さとCareCallへの反応を比較

- 固定効果:LTMあり・なし群
 - ランダム効果:参加者
 - 従属変数:コードの出現回数

- LTMあり群において、追加の分析

- 固定効果:LTMイベントの累積数
 - ランダム効果:参加者
 - 従属変数:コードの出現回数

Data Analysis Qualitative analysis

- 主題分析 (Thematic Analysis)
 - コールログとインタビューの記録を分析し、量的分析から得られた知見を補完する形でユーザーの反応やLTMに対する認識を詳細に描写

FINDINGS

*** $p < .001$; ** $p < .01$; * $p < .05$

Code	<i>LTM^{yes}</i> (# of calls = 576)	<i>LTM^{no}</i> (# of calls = 676)	95% CIs of Code Counts/Call
Meals-simple***	573 (99.4%)	455 (67.3%)	
Meals-detail***	101 (17.5%)	29 (4.3%)	
Sleep-simple***	192 (33.3%)	338 (50%)	
Sleep-detail	63 (10.9%)	53 (7.8%)	
Health-simple*	566 (98.3%)	567 (83.4%)	
Health-detail***	342 (59.4%)	182 (26.9%)	
Health-attitude*	80 (13.9%)	46 (6.8%)	
Clinical-simple***	488 (84.7%)	307 (45.4%)	
Clinical-detail***	183 (31.8%)	91 (13.5%)	
Activity-simple	396 (68.8%)	554 (82.0%)	
Social-detail	95 (16.5%)	74 (11.0%)	
Physical-detail*	67 (11.6%)	151 (22.3%)	
Wellbeing-simple	186 (32.3%)	237 (35.1%)	
Wellbeing-detail	9 (1.5%)	23 (3.4%)	
Hobbies	19 (3.3%)	44 (6.5%)	

FINDINGS

- LTMあり群はなし群と比べて健康に関してより深く自己開示した
 - 健康に関する自己開示はLTMの実装に関して最重要項目
 - 食事に関しても同様
- 一方、すべてのLTMトピックが自己開示を促進したわけではなかった
 - LTMのトピックに含まれている「訪問した場所」と「ペット」
 - Activity-simple、Social-detail、Hobbiesに関して群間に差がなかった
 - LTMのトピックに含まれている「睡眠」
 - LTMあり群のほうがSleep-simpleで自己開示が浅かった

FINDINGS

- LTMの設計が自己開示の頻度に影響を与えたことが示唆された
 - LTMトピックの優先順位
 - 1.健康 2.食事と睡眠 3.訪問した場所とペット

FINDINGS

Table 3: Frequency of codes and 95% Confidence intervals of code counts per call in the LTM^{yes} (colored bars) and LTM^{no} groups, which are relevant to reactions to the chatbot. The percentages do not add up to 100% as we multi-coded the dialogues.

Code	LTM^{yes} (# of calls = 576)	LTM^{no} (# of calls = 676)	95% CIs of Code Counts/Call
			0 0.25 0.5 0.75 1.0 1.25
Appreciation***	505 (87.7%)	337 (49.9%)	
Anthromorphization	59 (10.2%)	58 (8.6%)	
Negative feedback	20 (3.5%)	11 (1.6%)	
Disregard	139 (24.1%)	151 (22.3%)	
			0 0.25 0.5 0.75 1.0 1.25

FINDINGS

- Appreciation
 - チャットボットへの感謝 (Appreciation) を有意に多く表した ($p < 0.001$)
- Anthromorphization
 - 有意差はなかったものの、インタビューにおいて、いくつかのLTMイベントでユーザーは「チャットボットに人間らしさを感じた」と報告した

FINDINGS

- ネガティブな反応

- LTMの体験が親しみを提供することで、チャットボットに対するネガティブな反応を軽減することもわかった
 - 最初はネガティブな反応を示していたものの、LTMイベントを経験した後にはポジティブな反応に変わったケースがあった

FINDINGS

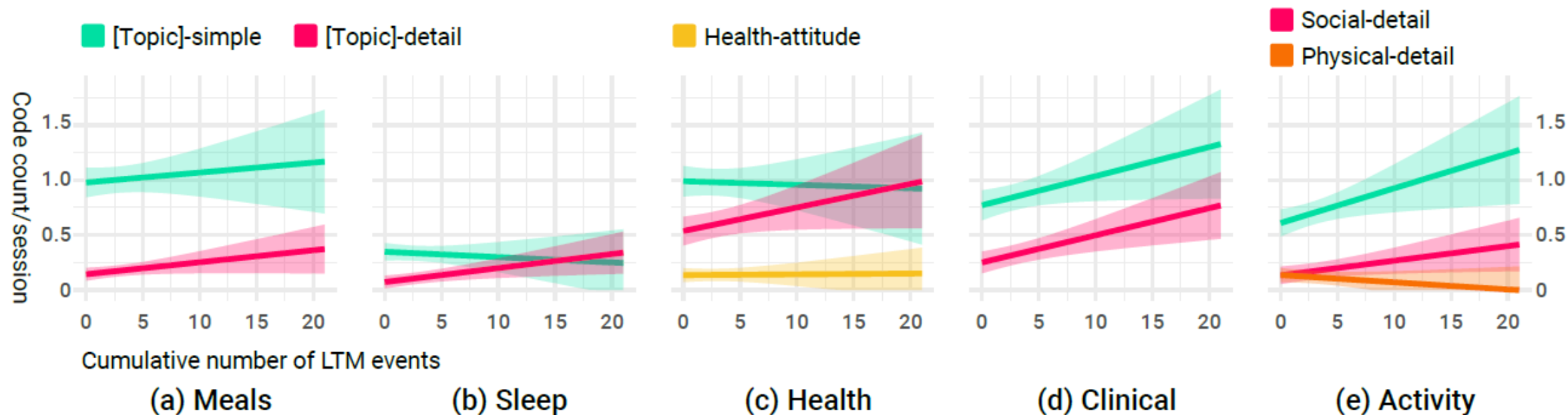


Figure 3: Estimated means and 95% confidence intervals of code counts about *Meals*, *Sleep*, *Health*, *Clinical*, and *Activity* by the cumulative number of LTM events in the *LTM^{yes}* group. The colored lines indicate the estimated means and the shaded areas indicate 95% confidence intervals of the code counts per call for each code. Overall, the repeated experiences of LTM events led to greater disclosure of more detailed information across the five categories.

FINDINGS

- LTMイベントの累積回数が「-detail」コードの合計数に正の影響を与えた
 - つまり、LTMあり群の参加者は、LTMイベントを経験するごとに、より多くの詳細情報を開示した($p < 0.001$)
 - 一方、「-simple」コードの合計数には有意な影響は見られなかった($p = 0.19$)。

FINDINGS

- 詳細な分析
 - Sleep-detail: $p = 0.01$
 - Clinical-detail: $p = 0.02$
 - Clinical-simple: $p = 0.047$
 - Activity-simple: $p = 0.02$
 - Social-detail: $p = 0.04$

FINDINGS

- LTMの自己開示促進に関する課題
 - 同じ内容を繰り返し尋ねる
 - ユーザーは新しい情報を開示する意欲を失い、簡潔な反応に終始するケースが見られた
 - LTMがトリガーする慢性的な健康問題に関する質問
 - 一部の参加者にとって問題が改善しない現実を思い出させることになり、個人的な健康に関するさらなる開示を抑制する要因となっていた
 - 健康に関するLTMトリガーの質問がプライバシーに対する懸念
 - ユーザーの自己開示に対する快適さに影響を与える可能性があることも示された

DISCUSSION

- 本研究の結果から
 - LTMがチャットボットを通じて自己開示を促進する上で有益であり、公衆衛生にとって価値ある機会を提供することが示された
 - LTMがユーザーに親しみを提供し、チャットボットとの関与を促進することも示された
 - 公衆衛生モニタリングにおける記憶の必要性和プライバシー保護との間の緊張を考慮する必要性が示唆された

DISCUSSION

- 特に健康に関しての自己開示が促進された
 - 主にLTMイベントの会話がユーザーに親しみを提供し、思いやりを示したことで、チャットボットに対する印象が向上した
 - この発見は、公衆衛生モニタリングのためのチャットボットにLTMを組み込むことの潜在的な利点を示している

DISCUSSION

- LTMの設計(トピックの優先度)がユーザーの会話パターンに与える影響
 - 公衆衛生モニタリングに役立つ重要なトピックに会話を誘導するのに効果的であった
 - 会話の幅を狭めることにつながった可能性もあった
- 選択的記憶(Selective Memory)
 - チャットボットにどのトピックを記憶させるべきか慎重に決定する必要がある
 - 特に社会的孤立の文脈では、個人の生活の広範な側面を記憶することが、「理解され、気にかけている」と感じさせる効果をもたらす
 - ユーザーの趣味や興味を記憶するように設計した場合、会話が本来の目的から逸れてしまう可能性がある

DISCUSSION

- サンプルの偏りについて
 - 社会的に孤立した女性の経験に偏っている可能性がある（特にLMTあり群は、男性17人：女性49人）
 - 最尤法検定の結果、性別がほとんどのコードに有意な影響を与えなかった（例外は「Health-attitude」と「Appreciation」）ことが示されたが、女性への偏りがコールログ分析における定性的な発見に影響した可能性がある
 - 孤独な生活を送る男性は、LTMのシステムに対する認識や自己開示のパターンが異なるかもしれない