

スキンケアカウンセリングの評価における 多様な模擬対話生成のためのLLMクライアントの提案

宮良 美桜[†] 澤崎 夏希[‡] 遠藤 聡志[§]
琉球大学[†] 琉球大学[‡] 琉球大学[§]

1 はじめに

近年、個々人の肌質や生活習慣に適したスキンケア商品の選択が、肌の健康維持や美容目的において重要性を増している。しかし、専門的な知識が必要であることや、ECサイトでは販売員による対面の助言や実物の試用が困難であることから、多くの消費者が商品選びに不安や失敗を感じている。実際、MMD研究所の調査によれば、初めて購入するスキンケア商品をECサイトで選んだ消費者のうち54.4%が「肌質に合わない」「期待した使用感でない」などの購入失敗を経験していると報告されている[1]。

このような課題に対し、化粧品販売員などの専門家によるスキンケアカウンセリングは有効な解決策となりうる。しかし、対面相談に対して心理的なハードルを感じるユーザーも少なくなく、より気軽に助言を得られる方法への需要が高まっている。

こうした背景のもと、筆者らの先行研究では、スキンケアカウンセリングの導入部分、すなわちユーザーの肌状態や生活習慣に関する情報を引き出す「ピフォーカウンセリング」に着目し、これを大規模言語モデル(LLM)を用いて実現することを試みた。その結果、スキンケアカウンセラーとして振る舞うLLMと人間クライアントとの対話を通じて、カウンセリングに必要な共通質問項目の実施率や共感表現の適切さを評価し、一定の有効性が初期実験により示された[3]。

しかしながら、十分な数の対話データに基づく評価は実施されておらず、評価の妥当性には依然として課題が残されている。年代、肌質、肌悩みといった多様な特性をもつ人間クライアントとの対話データを収集することは困難であり、多様性を担保した上で効率的に対話データを収集・評価する手法が求められている。

そこで本研究では、クライアント役を模倣するLLM(クライアントLLM)を設計し、スキンケアカウンセラーLLMとの模擬対話を自動生成する枠組みを提案する。この枠組みにより、多様なユーザープロフィールに基づいた対話データを低コストかつ効率的に生成・収集することを可能とし、スキンケアカウンセラーLLMの評価手法として有効な代替手段となり得ることを示すことを目的とする。

2 基礎概念や専門的技術の概要

2.1 スキンケアカウンセリングについて

日本スキンケア協会によると、スキンケアカウンセリングとは、肌や心の状態を見極めて要望を汲み取り、適切なホームケアや食生活、生活習慣に関するアドバイスを行うことを目的としたものであるとされている[2]。そのため、心理療法や精神療法を指す「心理カウンセリング」とは異なり、悩みや問題を解決するためにクライアントの情報を収集することに重点を置いている。

スキンケアカウンセリングでは、一般的にカウンセリングシートが用いられる。このシートには、クライアントの肌状態、健康状態、生活習慣、スキンケアの手順などの情報が記載されており、これらの情報を基にカウンセリングが進められる。

2.2 関連研究

本研究は、筆者らの先行研究[3]を発展させたものである。先行研究では、スキンケアカウンセリングにおけるピフォーカウンセリングを、LLMによって実現する手法に焦点を当て、スキンケアカウンセラーの試作を行った。スキンケアカウンセラーLLMと人間のクライアントとの対話において、カウンセラーLLMが共通質問として設定された内容を適切に提示できるかという「質問の実施率」と、クライアントに対して共感的な応答ができてきているかという「共感表現」の2点について評価を行った。質問の実施率については対話ログの確認を通じて、共感表現については人間クライアントからの主観的評価を通じて検証を行った。その結果、LLMが肌悩みに関する共通質問を適切に提示し、クライアントが共感的な応答として受け取っていることが確認された。しかし、十分な数の対話データに基づく評価は行われておらず、評価の妥当性に課題が残されていた。

そこで本研究では、クライアント役を模倣するLLMを設計し試作したスキンケアカウンセラーと対話させることで、LLM同士による模擬対話を生成する手法を提案する。LLM同士の対話生成に関する研究としては、心理カウンセリングや教育支援、交渉訓練など、対話シナリオの自動生成やモデル評価の目的で用いられた先行研究が挙げられる。例えば、Qiuらの研究では、カウンセラーとクライアントの役割をLLMに割り当て、心理カウンセリングの模擬対話データを生成する枠組みが提案されている[6]。本研究では、Qiuらの手法を参考にしつつ、スキンケアカウンセリングに特化したクライアント役LLMの生成を試みる。具体的には、ユーザープロフィールに肌質や肌悩みといった項目を用いてクライアントの個性を定義し、先行研究で設計したカウンセラーLLMとの対話を模倣的に生成する。この手法により、多様な対話データを低コストで収集可能とし、従来は実対話収集に依存していた、スキンケアカウンセラーLLMの評価効率化を目指す。

3 提案手法

本研究では、大規模言語モデルの一つであるChatGPTを活用し、スキンケアカウンセリングにおけるクライアント発話を自動生成する手法を提案する。本手法におけるクライアントの表現方法の概要については、図1に示す。クライアントを表現する方法として「年代」、「肌質」、「肌悩み」を組み合わせて表現する方法を用いることで、クライアントの様々な特性を表現でき、クライアントLLMに多様性をもたせることができる。

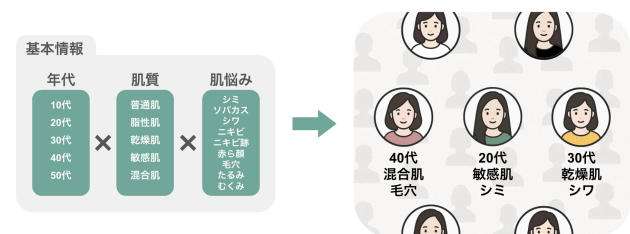


図1: クライアントの表現方法の概要

図2は提案手法の全体構成である。本手法は主に、「LLMによるカウンセリングシートの生成」と「クライアントLLMの作成と模擬対話の生成」の2段階から構成される。

本手法では、本来カウンセラーがスキンケアカウンセリングを行う

A Proposal for LLM-Driven Simulated Clients to Generate Diverse Dialogues for Evaluating Skin Care Counseling

[†] Mio Miyara, University of the Ryukyus

[‡] Natsuki Sawasaki, University of the Ryukyus

[§] Satoshi Endo, University of the Ryukyus

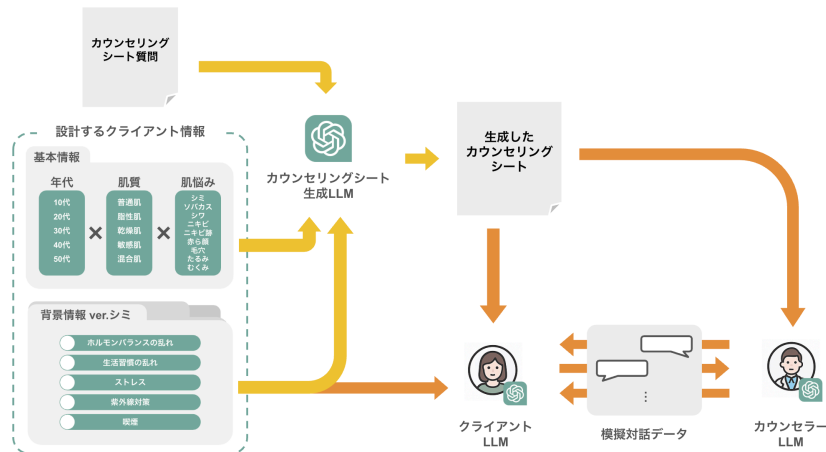


図2: 提案手法

際に使用される生成したカウンセリングシートを、カウンセリングだけではなく、LLM クライアントを生成するための材料として使用する。

3.1 LLM によるカウンセリングシートの生成

図2のLLMによるカウンセリングシートの生成では、クライアントの基本情報および背景情報に基づいたカウンセリングシートをLLMを用いて生成する。生成するカウンセリングシートの項目は、日本スキンケア協会の『しっかり学べるスキンケアカウンセリング教本』のカウンセリングシートを元に、個人情報の項目を除外して作成した。本段階においては、以下の情報を用いる。

1. 基本情報：年代・肌質・肌悩み
2. 背景情報：肌悩みごとに定義した追加情報
3. カウンセリングシートの質問

基本情報は、年代（5通り）、肌質（5通り）、肌悩み（9通り）の各カテゴリから1つずつ選択する構成とした。なお、肌質および肌悩みの選択肢は、カウンセリングシートに記載されている内容を元に設定している。

加えて、肌悩みに応じた背景情報を定義することで、カウンセリング内容の個別最適化を図る。この背景情報は、カウンセラーLLMが対話中に確認すべき肌悩みの原因に基づいて設計されており、各項目は「あり・なし」の2値で構成される。たとえば、肌悩みが「シミ」の場合、背景情報として11項目が定義されており、 $2^{11} = 2048$ 通りの組み合わせが存在する。したがって、肌悩みが「シミ」であるクライアントに対しては、年代・肌質・背景情報の組み合わせにより、 $5 \times 5 \times 2^{11} = 51,200$ 通りのバリエーションが存在する。このような構成により、複数の原因を柔軟に組み合わせたクライアントの実現が可能となる。

今回は、1つの肌悩みを持つクライアントを想定しているが、将来的に複数の肌悩みを持つケースを想定する場合には、すべての背景情報項目数を n とすると、理論上の組み合わせ数として $5 \times 5 \times 2^n$ 通りのクライアントパターンが想定可能である。たとえば、背景情報の総項目数が30であれば、 $5 \times 5 \times 2^{30} = 26,843,545,600$ 通りのクライアントを表現できることになる。

3.2 クライアントLLMの作成と模擬対話の生成

図2のクライアントLLMによる発話生成では、以下の情報を用いて模擬対話の生成を行う。

1. LLMによって生成されたカウンセリングシート
2. 背景情報：肌悩みごとに定義した追加情報
3. カウンセラーLLMの初回発話

カウンセリングシートの生成で使用した基本情報は、すでにカウンセリングシートに反映されているため本段階では使用しない。一方

で、背景情報はカウンセリングシートに全てが反映されるとは限らないため、発話生成においても利用する。

対話開始時の発話は、カウンセリングシート、背景情報、カウンセラーLLMの初回発話を入力として、クライアントLLMに生成させる。その後は、クライアントLLMの発話をカウンセラーLLMに入力し、次のカウンセラー発話を生成する。これを再びクライアントLLMに入力することで、次のクライアント発話を生成する。このプロセスを交互に繰り返すことで、クライアントLLMとカウンセラーLLMによる模擬対話を自動的に生成する。

4 実験と結果

本章では、提案手法の各段階における実験方法、得られた結果、およびそれに基づく考察について述べる。LLMによるカウンセリングシート生成およびクライアント役LLMにはchatGPTを、カウンセラー役には前研究[3]で試作したスキンケアカウンセラーLLMを用いた。

4.1 LLMによるカウンセリングシートの生成

4.1.1 プロンプト設計

表1に、プロンプトを示す。プロンプトでは、RolePromptingにより、LLMの役割を明確に示した。また、出力するカウンセリングシートへの回答が、実際の人の回答と差異がないように該当しない場合は空行にする指示や、記入形式（テキスト記入、選択肢選択など）についても明示的に指定した。

表1: LLMによるカウンセリングシート生成時のプロンプト例

<p>あなたは現在、スキンケアカウンセリングを受けに来たクライアントです。以下の基本情報と背景情報をもとにクライアントになりきり、提示するカウンセリングシートに質問文は表示せず回答のみで答えてください。該当しない質問は、空行にしてください。</p> <p># クライアントの情報 ## 基本情報 (略)</p> <p>## 背景情報 (略)</p> <p># カウンセリングシート 1. お名前を教えてください。 記入形式：テキスト 2. 生年月日を教えてください。 記入形式：日付 (略)</p>

4.1.2 結果と考察

カウンセリングシートの生成においては、基本情報として「年代：20代」、「肌質：混合肌」、「肌悩み：シミ」を設定した。加えて、背景情報として表2の「設定値」列に示す内容を用いた。なお、同表の「回答値」列については後述する。生成されたカウンセリングシートのうち、背景情報の影響がみられたと考えられる項目を表3に示す。

まず、生成結果の評価として、まずプロンプトに対する出力の忠実性を確認した。生成されたカウンセリングシートの項目のうち、背景情報が影響する項目を確認した結果、全体として出力は概ね指定通りであったが、背景情報として指定した「日焼け止めの使用」が適切に反映されないケースがあった。

次に、LLM クライアントによるシートと実際の人間の記入データを比較した。背景情報や基本情報に関連する設問の回答内容については、多くの項目で人間の回答と類似した傾向が見られた。一方で、「熟睡できたかどうか」や「外食の有無・頻度」など、プロンプトで具体的に指定されていない項目については、人間のクライアントと同様の回答が得られない場合が多い傾向が見られた。背景情報に関係する項目は概ね反映できているため、このあたりには一般的な回答を出力していると推測できる。人間の回答が一般的じゃなかったこともあり、そのような結果になったのではないかと考えられる。また、一部の項目（例：趣味、行っている運動の種類）においては、回答に多様性に乏しい傾向が見られ、画一的な回答傾向が見られた。

カウンセリングシート生成において、一部項目で同様の回答が得られなかったこと、似通った出力が多く見られる理由については、「代表性バイアス」が関係しているのではないかと考える。代表性バイアスとは、特定の集団や属性が訓練データ内で過剰または過少に表現されることで LLM の出力にバイアスが現れることである。Guo らの文献調査においても指摘されており、たとえば「科学やリーダーシップの文脈では男性が過剰に表現され、ケアに関する話題では女性が多く登場する」などが例として挙げられている [7]。

今回の日焼け止めの使用が適切に反映されなかったケースだと、一般的にはシミに悩んでいる場合、日焼け止めを日常的に使用するケースが多いことが理由だろう。このような傾向は訓練データにも多く含まれている可能性が高く、LLM がそのような「代表的（典型的）な」パターンを学習しており、プロンプトで「日焼け止めを使用していない」と明記されていたにもかかわらず、それに反する出力を行ったものと考えられる。

表 2: 生成時に設定したシミの背景情報とクライアント LLM の回答

項目	設定値	回答値
ホルモンバランスの乱れ	生理不順	あり
	妊娠・出産	なし
	閉経	なし
生活習慣の乱れ	睡眠不足	あり
	偏った食生活	あり
	運動習慣	あり
ストレス	あり	あり
紫外線対策	日焼け止めの使用	なし
	日傘の使用	あり
	帽子・UVカット衣類の使用	なし
喫煙	なし	なし
一致率	100% (11/11)	

4.2 LLM によるクライアントの発話生成

4.2.1 プロンプト設計

表 4 に、クライアントの発話生成に使用したプロンプト例を示す。本プロンプトでは、Role Prompting を用いて、LLM に対してクライアントとしての具体的な役割を明示的に与えている。カウンセリングシートおよび背景情報の内容は、想定されるクライアントの属性に応じて適宜変更する。また、カウンセラーの発話には、カウンセラー LLM に対し「スキンケアカウンセリングを始める」と入力した際の出

表 3: カウンセリングシート項目に対する LLM 出力の例

項目	クライアント 1	クライアント 2
名前	山田花子	佐藤美咲
生年月日	1998年5月15日	1997年6月15日
肌質	混合肌	混合肌
肌悩み	シミ	シミ
かぶれ	なし	なし
かぶれた化粧品		
皮膚科への通院	なし	なし
処方された薬		
使用化粧品	クレンジング、化粧水、乳液、 日焼け止め	洗顔料（敏感肌用）、化粧水（保湿タイプ）、乳液（乾燥肌用）
精神的な疲労	不安定	不安定
疲労の症状	ストレス	ストレス
睡眠時間	5	5
熟睡できたか	できない	できない
運動習慣（回/週）	2	1
運動内容	ヨガ、ジョギング	ウォーキング
普段の食事	不規則	不規則
外食の有無	あり（少ない）	あり（多い）
喫煙（本/日）	0	0

力を記入している。

表 4: LLM によるクライアント発話生成時のプロンプト例

あなたはスキンケアカウンセリングを受けに来たクライアントです。以下のカウンセリングシートと背景情報をもとに、クライアントになりきってカウンセラーと対話してください。
カウンセリングシート お名前：佐藤 美咲 生年月日：1997年6月15日 (略)
背景情報 (略)
カウンセラーの発話 カウンセラー：こんにちは、佐藤美咲さん。本日はスキンケアアドバイザーのためのカウンセリングをしていきますね。 カウンセラー：[肌質の確認]「混合肌」とご記入いただいておりますが、ご自身でもそのように感じていらっしゃいますか？ (略)

4.2.2 結果と評価

クライアント生成に使用した背景情報は表2の「設定値」列に示す内容を用いた。なお、同表の「LLM の回答」列については後述する。

カウンセリングシートは節 4.1.2 と同様に、基本情報として「年代：20代」、「肌質：混合肌」、「肌悩み：シミ」を、背景情報は表2の「設定値」列に示す内容を用いた。その中から、プロンプトに対する忠実性および人間との類似度が最も高いカウンセリングシートを用いた。生成されたクライアント LLM の発話の一部を表5に示す。

生成結果の評価として、プロンプトに対する出力の忠実性を確認した。その一環として、背景情報を考慮した発話生成が生成されているかを検証するため、カウンセラーとして各項目について有無を尋ねた。クライアント LLM の回答を表2の「回答値」列に示す。生成された発話はいずれも、背景情報として設定した内容と一致しており、背景情報を考慮した発話生成が可能であることが確認された。

生成した対話全体を確認したところ、プロンプトで設定した情報に基づき、おおむね適切な発話生成がされていた。特に、カウンセリングシートに関連する内容については、カウンセリングシートの入力情報を踏まえた受け答えがされていることを確認できた。一方で、一部

の項目に関しては、入力情報が十分に反映されていない場合も見られた。例えば、「日焼け止めの使用」に関する情報については、プロンプトに明記しているにもかかわらず、発話に反映されない場合があった。表 6 に反映されなかった例を、表 7 に反映されていた例をそれぞれ示す。表 6 の例では、「現在は日焼け止めを使用している」というような発話がされているが、プロンプトでは「日焼け止めを使用していない」と明示されている。これはプロンプト指定と矛盾しており、精度の課題を示している挙動である。このような誤りは、カウンセリングシートの生成時と同様に、肌悩みが「シミ」である場合には日焼け止めを使用しているだろうという LLM のバイアスによって、入力情報とは異なる内容が出力されている可能性がある。また、直前のカウンセラーの発話に注目すると、背景情報が反映されなかった場合には「これまでどのような対策を試してきましたか」といった曖昧な質問がなされているのに対し、背景情報が反映された場合には紫外線対策に関する具体的な質問が行われていた。このことから、カウンセラー側の質問の具体性が、クライアント側の応答に影響を与える可能性が示唆された。

結果として、カウンセラーとの対話は適切に行えるものの、プロンプトに忠実な発話を常に生成できるとは限らない点から、現時点では実際のクライアントの代替としては不十分である。代替可能とするためには、プロンプトに沿った発話を安定的に生成できるよう、代表性バイアスへの対策が求められる。

表 5: LLM による対話の生成結果例

佐藤美咲: こんにちは、よろしくお願ひします。はい、そうですね。Tゾーンは割と脂っぽくて、頬や顎の辺りは乾燥しやすいので、混合肌だと感じています。特に季節の変わり目に乾燥がひどくなることが多いです。
カウンセラー: [回答の再確認と共感] Tゾーンは脂っぽく、頬や顎は乾燥しやすいですね。季節の変わり目に乾燥がひどくなるということも、混合肌の方によくあるお悩みです。
カウンセラー: [かぶれの確認] それでは次に、基礎化粧品で「かぶれたことはない」とのご記入ですが、こちらは間違いないでしょうか？
カウンセラー: [事実確認] 化粧品でかぶれた経験は、本当に一度もないということで大丈夫ですか？
佐藤美咲: はい、基礎化粧品でかぶれたことは一度もないです。今使っている洗顔料や化粧水、乳液も特に問題なく使っているんで、肌には合っていると思います。

表 6: LLM による対話生成時にプロンプト指示を反映できていない例

カウンセラー: [共感とねぎらい] お仕事が忙しくて外出が増えた時期だったんですね。紫外線対策ができなかったのは仕方ないこともありますし、それだけ頑張っていた証拠ですね。
カウンセラー: [お手入れ経験の確認] そのシミを防ぐために、これまでどのようなスキンケアや対策を試されてきましたか？
佐藤美咲: これまで特にシミ対策としては、日焼け止めを使ったことがなかったので、最近では 日焼け止めを使うようにしています 。でも、まだ使い始めたばかりで、あまり続けられていないんです。それ以外では、シミに効くと言われているビタミン C 配合の化粧水を使ったりしていますが、あまり効果を実感できていない感じです。

表 7: LLM による対話生成時にプロンプト指示を反映できている例

カウンセラー: それではまず、紫外線対策について見直してみましょう。普段はどのような対策をされていますか？
佐藤美咲: 紫外線対策については、 日焼け止めは使っていない んです。日傘を使うことはありますが、帽子や UV カットの衣類は特に使っていないですね。日焼け止めを塗り忘れてしまうことが多いので、意識的にできていないのが現状です。

5 結論と展望

本研究では、スキンケアカウンセリングシステムにおける評価の効率化に向けて、クライアント役を模倣する LLM を設計し、カウンセラー LLM との模擬対話を生成する手法を提案した。具体的には、年代・肌質・肌悩みによる基本情報と、各肌悩みに応じた背景情報を組み合わせたカウンセリングシートを生成し、それに基づいて発話を行うクライアント LLM を生成した。

実験の結果、生成されたカウンセリングシートは、基本情報や背景情報に基づき、プロンプトの指示に概ね従った内容となっていた。人が記入したカウンセリングシートとの比較においても、傾向として類似性が確認された。また、生成されたクライアント発話は、カウンセラーの発話に対して不自然さがなく、クライアントの個性や抱える問題が一定程度反映された応答が得られた。

一方で、いくつかの出力では、プロンプトで明示された指示内容が反映されない、あるいは LLM の一般的な知識に引きずられた応答が確認された。これは、LLM の持つ事前学習のバイアスが影響していると考えられ、クライアント役としての忠実性に課題が残る。特に、背景情報と矛盾する内容を含む発話は、クライアント LLM の制御精度を高める必要性を示している。

以上の結果から、現時点ではクライアント LLM は概ねクライアントとしての振る舞いが可能であるものの、プロンプトに従わない場合があるため、実際のクライアントの完全な代替には至らない。しかし、実際のクライアントを模倣する手段としては有望であると考えられる。

今後は、より忠実な模倣を可能にするため、代表性バイアスへの対策の検討、評価指標の定量化、模擬対話と実際の対話との比較評価といった観点から、さらなる改良および実証を進めていく予定である。

参考文献

- [1] 石黒花菜. "化粧品品の EC 購入とセルフ AI 診断に関する調査". MMD 研究所. 2023-6-25. https://mmdlabo.jp/investigation/detail_2212.html, (2024-4-30)
- [2] 一般社団法人 日本スキンケア協会. しっかり学べるスキンケアカウンセリング教本. 一般社団法人 日本スキンケア協会, 2014
- [3] 宮良 美桜, 遠藤 聡志, 澤崎 夏希. LLM を用いたスキンケアカウンセリングの試作と収集した情報の評価. 情報処理学会 第 87 回全国大会, 2025.
- [4] Van Buren D. 2023. Guided scenarios with simulated expert personae: a remarkable strategy to perform cognitive work. ArXiv:2306.03104.
- [5] Logan IV R, Balažević I, Wallace E, Petroni F, Singh S, Riedel S. 2022. Cutting down on prompts and parameters: simple few-shot learning with language models. Findings of the Association for Computational Linguistics: ACL 2022, 2824-2835.
- [6] Huachuan Qiu, Zhenzhong Lan. 2024. Interactive Agents: Simulating Counselor-Client Psychological Counseling via Role-Playing LLM-to-LLM Interactions. arXiv:2408.15787. <https://arxiv.org/abs/2408.15787>
- [7] Guo, Yufei and Guo, Muzhe and Su, Juntao and Yang, Zhou and Zhu, Mengqiu and Li, Hongfei and Qiu, Mengyang and Liu, Shuo Shuo. 2024. Bias in Large Language Models: Origin, Evaluation, and Mitigation. arXiv:2411.10915.