

多言語研究室環境における学生間交流支援のための ユーザの個性を聞き取る LLM 対話システムの試作

Prototyping an LLM-based User Persona Interview System for Supporting Student Communication in Multilingual Laboratory Environments

許 子恒^{1*}
Ziheng Xu¹

白松 俊¹
Shun Shiramatsu¹

長澤 史記¹
Fuminori Nagasawa¹

¹ 名古屋工業大学大学院 工学研究科

¹ Graduate School of Engineering, Nagoya Institute of Technology

Abstract: In multilingual laboratory environments, students often have difficulty initiating conversations and understanding each other because of language differences and limited mutual knowledge. This paper proposes an LLM-based conversational user information collection system for supporting student communication. The system collects users' interests, research topics, and personality-related information through dialogue and generates user cards that can be used as cues for topic finding and mutual understanding. We compared a structured baseline method with a proposed method that embeds information collection in a more natural dialogue flow. An experiment with 16 participants was conducted using questionnaires, Big Five estimation error based on BFI-10, and Discord communication log analysis. The results showed no significant advantage of the proposed method in subjective dialogue experience, while both methods received mean item ratings above the scale midpoint. The proposed method received significantly higher evaluations for user-card usefulness and communication support in the one-sided test based on the research hypothesis. It also yielded significantly lower overall personality estimation error relative to BFI-10. These results suggest that the proposed method may collect and organize user information useful for subsequent communication while maintaining a certain level of dialogue experience.

1 はじめに

近年、日本の大学研究室では、留学生や異なる言語的背景を持つ学生が同じ研究環境で活動する機会が増えている。このような多言語研究室環境では、言語的背景の違いや相互理解の不足により、相手の興味関心や研究内容を把握し、会話のきっかけを得ることが難しい場合がある。短時間の自己紹介では、その後の話題発見につながる具体的な情報を十分に共有しにくい一方、質問紙や入力フォームによる情報収集は、必要な情報を整理しやすいものの、回答が単調になりやすい。

大規模言語モデル (Large Language Model; LLM) の発展により、自然言語による柔軟な対話を通じてユーザ情報を収集し、その内容を要約・整理することが可

能になっている。LLM を用いた対話システムは、ユーザの発話内容に応じて追質問や話題展開を行えるため、興味関心、研究内容、性格傾向などからなるユーザの個性を対話的に収集できる可能性がある。しかし、LLM 対話によって収集した個性に関する情報をユーザカードとして整理し、多言語研究室環境における学生間交流支援へ接続する方法は、十分に検討されていない。

本研究では、多言語研究室環境における学生間交流支援を目的として、LLM を用いた対話型ユーザ情報収集システムを提案する。本システムは、対話を通じて興味関心、研究内容、パーソナリティ関連情報を収集し、話題発見と相互理解の手がかりとなるユーザカードを生成する。

本稿では、段階分離型の情報収集を行うベースライン手法と、情報収集を一連の対話に組み込んだ提案手法を比較する。対話体験、ユーザカードおよび交流支

*連絡先: 〒 466-8555 愛知県名古屋市中区御器所町
名古屋工業大学 2 号館 白松研究室 2 階 202B
E-mail: xuziheng@srmtlab.org

援に関する主観評価、BFI-10の結果を基準としたパーソナリティ推定誤差、ならびに Discord 上の交流ログから、後続の交流支援に利用可能なユーザ情報の収集方法を検討する。

本稿の構成は以下の通りである。第2章では、関連研究について述べる。第3章では、提案システムの構成について説明する。第4章では、実験条件を含む実験方法について述べる。第5章では、実験結果を示す。第6章では、結果に基づく考察を行い、第7章で本稿をまとめる。

2 関連研究

2.1 多言語研究室環境における学生間交流支援

多言語・異文化環境における学生間交流については、言語差や文化差が相互理解や関係形成に影響を与えることが指摘されている。Yu らは、香港の大学に在籍する国際学生を対象とした調査を通じて、国際学生が多様な背景を持つ学生との交流を望んでいるにもかかわらず、現地学生との実際の交流が限定的であることを報告している [1]。また、文化的・言語的差異が学生間の壁となり、交流が近い背景を持つ学生に偏る状況を示している。

高等教育における学生間交流を支援する手段として、オンラインチャットやチャットボットを活用する研究も行われている。Abbas らは、成人学生を対象に、オンラインチャットプラットフォームとチャットボットを用いた学生主導コミュニティを試行し、オンライン交流への参加を支える要因と運用上の課題を報告している [2]。これらの研究は交流の場を形成する重要性を扱う一方、交流に先立って相手の興味関心や研究内容を整理し、話題発見と相互理解を支援する方法は十分に扱っていない。

2.2 Big Five および BFI-10 に基づくパーソナリティ評価

本研究では、ユーザの個性を表す情報の一部として、Big Five に基づくパーソナリティ傾向を扱う。Big Five は、外向性、協調性、誠実性、神経症傾向、開放性の5次元から性格傾向を捉える枠組みである。Rammstedt と John は、回答時間が制限される場面で Big Five を短時間に測定するため、BFI-44 を基にした 10 項目版尺度である BFI-10 を提案した [5]。Carciofo らは、中国語版 BFI-44 および BFI-10 を心理測定学的に評価し、中国語による Big Five 研究での利用可能性を報告している [6]。

BFI-10 は各次元が2項目で構成される短縮尺度であり、詳細な性格傾向の測定には限界がある。そこで本研究では、BFI-10 の結果を絶対的な正解とはみなさず、LLM 対話システムによる推定結果を比較するための基準として用いる。なお、提案手法では対話中の質問候補に BFI-2 に基づく質問データベースを用いるのに対し、BFI-10 は実験時の評価基準として用いる。

2.3 LLM対話によるユーザ理解とパーソナリティ推定

近年、大規模言語モデル (Large Language Model; LLM) は、自然言語による柔軟な対話を可能にするだけでなく、ユーザの発話内容から心理的特徴や個人差を推定する手段としても注目されている。Peters らは、GPT-4 を用いたチャットボットによる自由形式の対話から Big Five を推定し、パーソナリティ関連情報を引き出す条件で推定結果と自己報告値との対応が高くなる一方、対話体験が必ずしも低下しなかったことを報告している [3]。また、Matšenas らは、ガイド付き LLM 対話から推定された Big Five の各次元スコアを IPIP-50 の結果と比較し、特性による差はあるものの、一定の対応関係が得られたと報告している [4]。

これらの研究は LLM 対話によるパーソナリティ推定の可能性を検討しているが、得られた性格傾向に興味関心や研究内容と組み合わせ、学生間交流支援に用いる点は対象としていない。本研究では、パーソナリティ関連情報をユーザの個性の一部として扱い、ユーザカードによる話題発見と相互理解に接続する。

以上より、本研究は、多言語研究室環境を対象に、LLM が興味関心、研究内容、性格傾向からなるユーザの個性を対話的に収集し、ユーザカードとして提示することで学生間交流を支援する点に位置付けられる。

3 提案システム

本章では、本研究で構築した LLM を用いた対話型ユーザ情報収集システムについて述べる。本システムは、ユーザの興味関心、研究内容およびパーソナリティ関連情報を対話的に収集し、学生間交流に利用するユーザカードを生成する。ユーザカードは、交流前に相手理解、共通点の発見、および話題発見の手がかりを提示することで、学生間交流を支援する。

3.1 システムの全体構成

ユーザは、ユーザ名、国籍、使用可能な言語などの基本情報を入力した後、AI エージェントと対話する。

システムは対話内容を分析してユーザカードを生成・表示する。本システムの機能範囲はユーザカードの生成・表示までであり、カードを用いた交流はDiscordなどの外部コミュニケーションツール上で行う。

3.2 収集するユーザ情報

本システムでは、GPT-5.2を用い、基本情報、興味関心、研究内容、パーソナリティ関連情報を収集する。基本情報にはユーザ名、国籍、使用可能な言語を含む。興味関心には趣味や最近関心を持つ話題、研究内容には研究テーマ、目的、手法、課題を含む。パーソナリティ関連情報には、Big Fiveの外向性、協調性、誠実性、神経症傾向、開放性に関する性格傾向を用いる。収集した情報は対話内容に基づいて要約し、ユーザカードの各項目として提示する。

3.3 対話におけるパーソナリティ関連質問の統合

提案手法は、興味関心から研究内容へ展開する一連の対話に、パーソナリティ関連情報の収集を組み込む方式である。質問候補には、BFI-2に基づく質問データベースを用いる。本実装では、興味関心に関する対話に開放性および外向性、研究内容に関する対話に協調性、誠実性および神経症傾向に対応する質問を配置した。各質問は、対象となる次元の質問候補からデータ検索を支援するフレームワークであるLlamaIndex[7]を用いて現在の対話内容に基づいて検索され、文脈に合う表現へ書き換えられる。

回答後、GPTが性格傾向のスコアリングに利用できるかを0.0から1.0の範囲の信頼度として判定し、0.6以上の場合にスコアを付与する。0.6未満の場合は、同じ次元と方向に対応する別の質問を提示する。別の質問による確認を3回繰り返しても基準を満たさない場合は、それまでの対話全体に基づいて当該次元と方向のスコアを推定する。これらの処理は、性格傾向を単独で診断するためではなく、興味関心や研究内容と関連付けてユーザの特徴を整理し、ユーザカード上で交流の手がかりとして提示するために用いる。

3.4 ユーザカードの生成と交流支援への利用

対話終了後、システムは収集した情報から、公開情報とユーザ本人向け情報で構成されるユーザカードを生成する。公開情報には、基本情報、興味関心のタグ、研究テーマ、研究内容の説明、ユーザの特徴を表す一文を含み、他の学生が相手を理解し、共通点や話題を見つけるために用いる。ユーザ本人向け情報には、交流時

のアドバイスとパーソナリティ傾向のレーダーチャートを含むが、実験では公開情報のみをスクリーンショットとしてDiscord上で共有した。

4 実験方法

4.1 実験目的

本実験の目的は、情報収集方式の違いが、対話体験、ユーザカードの有用性、パーソナリティ推定誤差、およびユーザカード使用前後の交流に与える影響を検討することである。本研究では、提案手法により、ユーザカードおよび交流支援に関する評価が高くなり、BFI-10を基準とした推定誤差が小さくなると仮定した。一方、対話体験については方向性を定めずに確認した。

4.2 実験参加者

実験参加者は、同一研究室に所属する学生16名であり、日本、中国、バングラデシュ、ブルキナファソなど、複数の国籍・言語背景を持つ学生が含まれていた。参加者を4名ずつグループAからグループDの4グループに分け、全員がベースライン手法と提案手法の両方を体験した。分割は無作為割付ではなく、実験日程、参加可能時間、国籍・言語背景の混合、および条件の使用順序の均衡を考慮して行った。

実験前に、実験目的、取得するデータの種類の種類、個人情報の扱い、取得データを研究目的にのみ利用すること、および個人を特定できる形で外部に公開しないことを参加者に説明した。

4.3 実験条件

本実験では、ユーザカードを生成するための情報収集方式の違いを検討するため、ベースライン手法と提案手法の2条件を設定した。両条件では、同一のLLMであるGPT-5.2を用い、収集する情報の種類およびユーザカードの出力様式も同一とした。両条件の主な違いは、パーソナリティ関連情報の収集方法である。

ベースライン手法は、興味関心、研究内容、パーソナリティ関連情報を個別の段階で収集する段階分離型の方式である。興味関心と研究内容について各4ターンの対話を行い、ユーザが手動で段階を切り替えた後、Big Fiveに対応するパーソナリティ関連質問を提示する。一方、提案手法では、興味関心や研究内容に関する一連の対話の中にパーソナリティ関連質問を組み込み、現在の対話内容に基づいて質問候補を検索し、文脈に合う表現へ書き換えて提示する。また、回答が十分でない場合には、信頼度に基づいて追加確認を行う。

表 1: グループごとの参加者数と条件の使用順序

グループ	参加者数	使用順序
A	4	提案手法 → ベースライン手法
B	4	ベースライン手法 → 提案手法
C	4	提案手法 → ベースライン手法
D	4	ベースライン手法 → 提案手法

したがって、本実験で比較する差分は、ユーザカードの表示形式ではなく、ユーザカードを生成するための情報収集過程、特にパーソナリティ関連質問の統合方法である。参加者はログイン時に使用可能な言語を入力し、各自が使いやすい言語で実験用システムと対話した。本実験は、同一参加者が2条件を体験する対応ありの実験計画とした。使用順序の影響を軽減するため、条件の使用順序をグループ間で分けた。表1に、グループごとの参加者数と条件の使用順序を示す。参加者には、2種類のシステムを体験することのみを説明し、各システムがベースライン手法または提案手法のいずれであるかは明示しなかった。実験は2週間以内の複数日に分けて実施した。

4.4 実験手順

実験開始前に、参加者は各自で BFI-10 に回答した。BFI-10 は5件法で回答し、パーソナリティ推定結果の比較基準として用いた。

各グループは、まずシステム使用前に Discord 上で10分間交流した。次に、1つ目の条件で実験用システムと対話し、対話終了後にアンケート A へ回答した。生成されたユーザカードの公開情報をスクリーンショットとして共有し、他の参加者のカードを閲覧した後、Discord 上で10分間交流してアンケート B へ回答した。続いて2つ目の条件でも同じ手順を実施した。このため、各グループから、システム使用前、1つ目の条件の使用後、2つ目の条件の使用後の計3回の Discord 交流ログを取得した。

4.5 評価指標

アンケート A は、実験用システムとの対話体験を評価するための質問紙であり、8項目から構成される。対話の自然さ、応答の適切さ、話題切り替え、負担感、情報伝達の効率、対話長などを評価した。アンケート B は、生成されたユーザカードおよび学生間交流支援に関する主観評価を行う質問紙であり、10項目から構成される。ユーザカードの分かりやすさ、自己理解、相互理解、共通点や話題の発見、学生間交流への有用性、今後の利用意向などを評価した。両質問紙はいずれも

7件法で回答し、値が高いほど肯定的な評価を表す。反転項目は含めず、各参加者について、条件ごとにアンケート A の8項目およびアンケート B の10項目の算術平均をそれぞれの総合得点とした。全参加者が両条件の全項目に回答しており、欠損値はなかった。

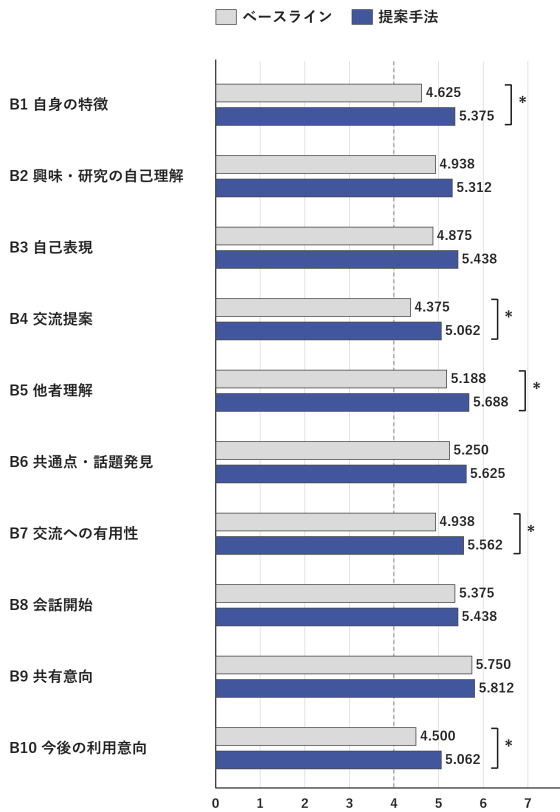
パーソナリティ推定については、実験前に回答した BFI-10 の結果を基準とし、各手法によって推定された Big Five の各次元スコアとの誤差を評価した。BFI-10 およびシステム推定結果はいずれも1から5の範囲で扱った。評価には平均絶対誤差 (Mean Absolute Error; MAE) を用い、Big Five 全体の MAE (以下、全体 MAE) と、外向性、協調性、誠実性、神経症傾向、開放性の各次元における MAE を算出した。

Discord 交流ログについては、ユーザカード使用前後の交流状況を補足的に分析した。分析対象は、各グループにおける3回の10分間交流ログである。総発言数は各交流における全発言数とした。メンション付き発言数は、Discord 上で特定の参加者を明示的にメンション (@) した発言の総数とし、「全体」に向けた発言は含めなかった。話題数は発言内容に基づいて手作業で分類した話題の数とした。ペア交流数は、参加者2名の組合せごとに、一方が他方を明示的にメンション (@) した発言を、発言方向を区別せず合計した値とした。交流のあったペア数は、ペア交流数が1以上であった参加者2名の組合せの数とした。

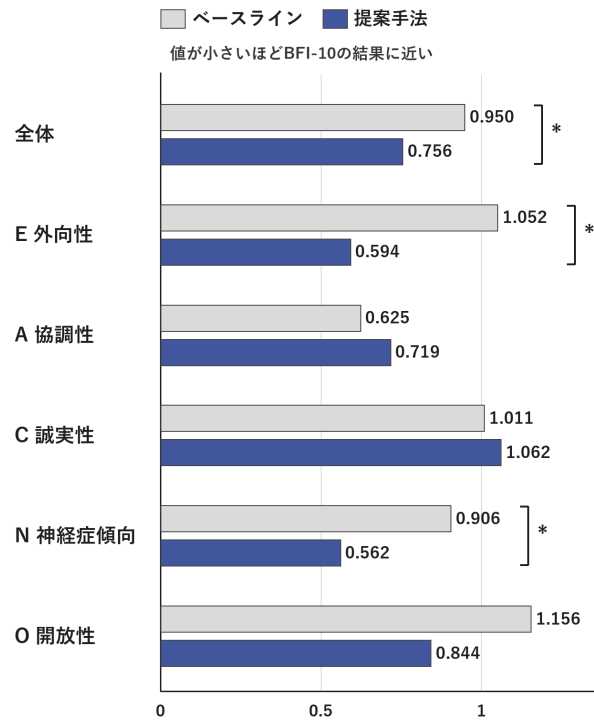
4.6 分析方法

質問紙評価およびパーソナリティ推定誤差の比較には、ウィルコクソンの符号付順位検定を用いた。本実験は同一参加者による対応ありデータであり、質問紙評価には Likert 尺度を用いたためである。アンケート A、アンケート B、全体 MAE および各次元の MAE について、ベースライン手法と提案手法を比較した。差が0となるペアは順位和の計算から除外したため、検定に用いる有効標本数は分析項目により異なる。効果量 r は、検定統計量から得られる Z 値を用いて $r = |Z|/\sqrt{N}$ とし算出し、 N には差が0のペアを除いた有効標本数を用いた。

アンケート A については、提案手法が対話体験を改善する方向を事前に仮定せず、両側検定を用いた。アンケート B については、提案手法によりユーザカードおよび交流支援に関する評価が高くなるという仮説に基づき、提案手法がベースライン手法より高い方向の片側検定を用いた。全体 MAE および各次元の MAE については、提案手法により BFI-10 を基準とした推定誤差が小さくなるという仮説に基づき、提案手法がベースライン手法より小さい方向の片側検定を用いた。Discord 交流ログについては、システム使用前、1つ目



(a) アンケート B



(b) BFI-10 基準の全体 MAE および次元別 MAE

図 1: (a) アンケート B の 10 項目の平均値, (b) BFI-10 の結果を基準とした全体 MAE および次元別 MAE。ベースライン手法を灰色, 提案手法を青色で示す。括弧は両手法の対応のある比較を表し, アスタリスクは研究仮説に基づく片側のウィルコクソンの符号付順位検定による 5%水準の有意差を示す ($p < .05$)。 (a) の破線は 7 件法の間値 4 を示し, (b) は値が小さいほど BFI-10 の結果に近いことを表す。

のシステム使用後, 2 つ目のシステム使用後の変化を記述的に分析した。使用順序の影響が含まれる可能性を考慮し, 交流ログ分析は主たる検定結果を補足する分析として扱った。

5 実験結果

5.1 対話体験に関する評価

アンケート A の総合得点は, ベースライン手法で 4.984, 提案手法で 4.805 であった。ウィルコクソンの符号付順位検定の結果, 両条件間に有意差は認められなかった (両側検定, $p = .338$, $r = .239$)。表 2 に, アンケート A, アンケート B および全体 MAE の総合結果を示す。

項目別の比較でも, 8 項目のいずれにも有意差は認められなかった。両条件とも全項目の平均値が 7 件法の間値である 4 を上回ったが, 対話体験に関して提案手法の有意な優位性は確認されなかった。

表 2: 主要評価指標の総合結果

指標	ベースライン手法	提案手法	n	p	r
アンケート A	4.984	4.805	16	.338	.239
アンケート B	4.981	5.438	14	.030*	.503
全体 MAE	0.950	0.756	13	.009*	.659

アンケート A および B は 7 件法の項目平均による総合得点である。全体 MAE は値が小さいほど BFI-10 の結果に近いことを表す。有効 n は条件間の差が 0 であるペアを除いた数である。アンケート A の p 値は両側検定, アンケート B および全体 MAE の p 値は研究仮説に基づく片側検定による。* は 5%水準の有意差を示す。

5.2 ユーザカードおよび交流支援に関する評価

アンケート B の総合得点は、ベースライン手法で 4.981、提案手法で 5.438 であった。研究仮説に基づく片側検定の結果、提案手法はベースライン手法より有意に高かった ($p = .030$, $r = .503$)。

図 1(a) に 10 項目の平均値を示す。項目別でも、自身の特徴の整理 (B1: ベースライン手法 4.625, 提案手法 5.375, $p = .007$, $r = .679$)、交流提案の参考性 (B4: ベースライン手法 4.375, 提案手法 5.063, $p = .038$, $r = .548$)、他者理解 (B5: ベースライン手法 5.188, 提案手法 5.688, $p = .044$, $r = .483$)、研究室内交流への有用性 (B7: ベースライン手法 4.938, 提案手法 5.563, $p = .013$, $r = .767$)、および今後の利用意向 (B10: ベースライン手法 4.500, 提案手法 5.063, $p = .017$, $r = .718$) において、提案手法が有意に高かった。共通点・話題発見 (B6)、会話開始 (B8)、共有意向 (B9) などでは、有意差は認められなかった。

5.3 パーソナリティ推定誤差

BFI-10 の結果を基準として、各条件で推定された Big Five の各次元スコアとの MAE を比較した。MAE が小さいほど、推定結果が BFI-10 の結果に近いことを表す。全体 MAE はベースライン手法で 0.950、提案手法で 0.756 であり、研究仮説に基づく片側検定の結果、提案手法で有意に小さかった ($p = .009$, $r = .659$)。

図 1(b) に BFI-10 基準の全体 MAE および次元別 MAE を示す。次元別では、外向性においてベースライン手法が 1.052、提案手法が 0.594 であり ($p = .002$, $r = .858$)、神経症傾向においてベースライン手法が 0.906、提案手法が 0.563 であり ($p = .019$, $r = .577$)、いずれも提案手法の MAE が有意に小さかった。開放性では提案手法の MAE が数値上小さかったものの、有意差は認められなかった。協調性および誠実性でも有意差は認められず、両次元では提案手法の平均 MAE がベースライン手法よりわずかに大きかった。

5.4 Discord 交流ログの分析

各グループについて、システム使用前と初回条件使用後に取得した 10 分間の Discord 交流ログを記述的に分析した。図 2 に、総発言数、メンション付き発言数および話題数について、初回使用後から使用前を差し引いた平均変化量を示す。

提案手法を最初に使用したグループ A および C では、総発言数とメンション付き発言数の平均変化量はいずれも 5.5 件、話題数は 1.0 件であった。ベースライ

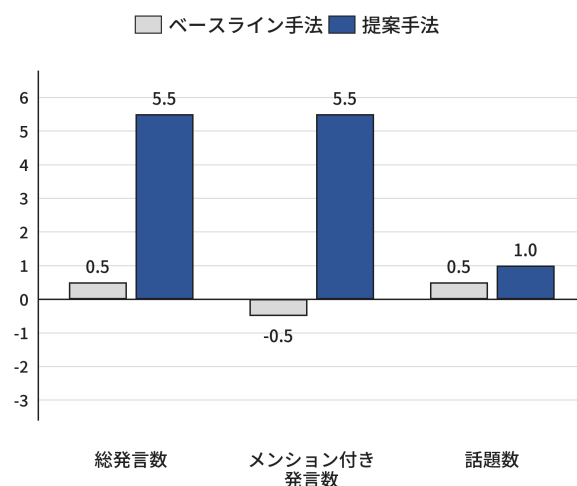


図 2: 初回条件使用前後における Discord 交流指標の平均変化量。各値は初回使用後から使用前を差し引き、提案手法はグループ A および C、ベースライン手法はグループ B および D の平均を示す。本結果はグループ単位の記述的比較である。

ン手法を最初に使用したグループ B および D では、総発言数が 0.5 件、メンション付き発言数が -0.5 件、話題数が 0.5 件であった。したがって、提案手法を最初に使用した 2 グループでは、総発言数およびメンション付き発言数の増加量が大きかった。ただし、各手法の平均値は 2 グループのみから算出しており、ベースライン手法ではグループ B で増加、グループ D で減少するなど、グループ間で変化の方向が異なった。

さらに、参加者 2 名の組合せごとにペア交流数を集計した。提案手法を最初に使用したグループ A では、交流のあったペア数が 2 組から 5 組、ペア交流数の合計が 5 件から 10 件となった。グループ C では、交流のあったペア数は 3 組のままであったが、ペア交流数の合計は 6 件から 12 件となった。一方、ベースライン手法を最初に使用したグループ B でも、交流のあったペア数が 3 組から 5 組に増加した。グループ D では、交流のあったペア数は 2 組のままであり、ペア交流数の合計は 8 件から 5 件に減少した。

ただし、これらは異なるグループを単位とした記述的比較であり、システム使用順序、グループ構成および参加者間の既存関係の影響を含む。そのため、本分析は交流促進効果を統計的に示すものではなく、質問紙結果を補足する結果として扱う。

6 考察

6.1 対話体験と情報収集方式

アンケート A では提案手法の明確な優位性は確認されなかったが、両条件とも全項目の平均値が 7 件法の間値を上回り、一定水準の対話体験を示した。

提案手法は情報収集を一連の対話に統合したが、アンケート A の自由記述では、AI 発言の長さ、複数質問の同時提示、質問間の関連性および話題切替に関する問題が指摘された。これらが回答のしやすさに影響し、融合的な情報収集の利点を相殺した可能性がある。一方、ベースライン手法は各段階の目的が明確であり、構造的な分かりやすさが一定の評価につながったと考えられる。また、両手法は同一の LLM を用いて同じ種類のユーザ情報を収集しており、共通する応答品質も対話体験の差を小さくした可能性がある。

6.2 ユーザカードとパーソナリティ関連情報

アンケート B では、研究仮説に基づく片側検定において、提案手法の総合得点がベースライン手法より有意に高かった。ただし、すべての項目で有意差が認められたわけではなく、ユーザカードおよび交流支援に関する評価が全面的に向上したとはいえない。

自身の特徴の整理、交流提案の参考性、他者理解、研究室交流への有用性、および今後の利用意向という個別項目でも、提案手法で有意に高い評価が見られた。興味関心、研究内容および性格傾向を同一の対話文脈で収集し、必要に応じて追加質問を行ったことで、ユーザの個性を一貫した情報として整理しやすくなり、カードの有用性や利用意向につながった可能性がある。一方、共通点・話題発見、会話開始および共有意向は両条件で比較的高く、明確な差は認められなかったため、共通形式のユーザカード自体も交流の手がかりとなった可能性がある。両条件でカード形式が共通していることから、評価の差は提示形式よりも、対話を通じて収集・整理された情報の内容に関係した可能性がある。

BFI-10 の結果を基準とした全体 MAE は提案手法で有意に小さく、全体として BFI-10 の結果により近いパーソナリティ関連情報が得られた。外向性および神経症傾向でも MAE が有意に小さく、具体的な経験や状況と結び付けた質問と追加確認が、推定に利用可能な情報の収集に寄与した可能性がある。

ただし、提案手法がすべての次元で BFI-10 の結果に近づいたわけではなく、推定の有用性は性格次元によって異なった。誠実性では、研究内容や課題への取り組みを語る実験場面での真面目な応答が、一般的な性格傾向として強く解釈された可能性がある。開放性についても、興味関心に関する発言の豊富さや趣味の

多様さは、BFI-10 で扱う想像力や芸術的関心と必ずしも一致しない。本実装では、開放性と外向性を興味関心に関する対話、協調性、誠実性および神経症傾向を研究内容に関する対話に配置したため、次元別の推定誤差には性格次元の特徴だけでなく、質問を提示した話題や文脈も影響した可能性がある。ただし、本実験では異なる配置や提示順序を比較しておらず、それぞれの影響を区別することはできない。

6.3 交流ログから得られた示唆

提案手法を最初に使用したグループでは、総発言数およびメンション付き発言数の平均増加量がベースライン手法を最初に使用したグループより大きかった。ユーザカードに示された興味関心や研究内容が、特定の参加者へ話しかける手がかりとなった可能性がある。ただし、各手法は 2 グループずつであり、ベースライン手法でもグループ B では増加が見られた一方、グループ D では減少したため、平均値のみから手法の効果を断定することはできない。ペア単位では、グループ A で交流のあったペア数が増加し、グループ C ではペア数を維持したままペア交流数の合計が増加した。このことは、ユーザカードが交流相手を広げる手がかりとなる場合と、既存の参加者間で発言を展開する手がかりとなる場合の双方がある可能性を示している。一方、グループ B でも交流のあったペア数が増加しており、使用順序、グループ構成および既存関係の影響を分離できないため、交流促進効果を示す結果ではない。話題数も両手法で平均的に増加したが、その差は小さく、話題数の増加は提案手法に固有の変化ではなかった。

6.4 制限事項と今後の課題

本研究の参加者は同一研究室の 16 名に限られ、参加者間の既存関係を事前に測定していない。また、グループ分けは無作為ではなく、使用順序とグループ構成の影響を完全には排除できない。Discord ログも各 10 分間のグループ単位データである。さらに、アンケート A および B は主観評価であり、BFI-10 も短縮尺度を比較基準として用いたものである。ユーザカードを用いない条件および長期観察を設けていないため、カード自体の効果や交流関係の持続的変化は検証できていない。

今後は、一度に提示する質問数、AI 発言の長さおよび話題遷移を調整するとともに、各次元の情報充足度と対話文脈に基づいて質問を動的に選択する方法を検討する必要がある。また、参加者と対象研究室を拡大し、事前関係と使用順序を統制した上で、ユーザカー

ドを用いない条件との比較および長期的な交流観察を行う必要がある。

7 おわりに

本研究では、多言語研究室環境における学生間交流支援を目的として、LLM 対話を通じてユーザの興味関心、研究内容およびパーソナリティ関連情報を収集し、話題発見と相互理解の手がかりとなるユーザカードを生成するシステムを提案した。実験では、段階分離型のベースライン手法と、情報収集を一連の対話に組み込む提案手法を比較した。

提案手法は主観的な対話体験において明確な優位性を示さなかったが、両手法とも一定水準の対話体験を示した。一方、提案手法では、ユーザカードおよび交流支援に関する総合評価が有意に高く、個別項目でも自身の特徴の整理、交流提案の参考性、他者理解、研究室内交流への有用性および今後の利用意向が有意に高かった。また、BFI-10 の結果を基準とした全体 MAE、外向性および神経症傾向の MAE が有意に小さかった。Discord ログからも、ユーザカードが交流対象の拡大や既存交流の展開に利用される可能性が記述的に示された。以上から、提案手法はすべての評価でベースライン手法を上回るものではないが、一定水準の対話体験を維持しながら、学生間交流支援に用いるユーザの個性をユーザカードとして整理する情報収集方法としての可能性を示した。

今後は、各性格次元の情報充足度を考慮しながら、対話文脈に応じて質問を動的に選択する方法を検討するとともに、一度に提示する質問数、AI 発言の長さおよび話題遷移を改善し、質問統合の自然さとユーザ体験の向上を図る。また、現行システムと生成されたユーザカードを基盤として、交流時の助言提示、対話相手の推薦・マッチング、および話題推薦などの学生間交流支援機能を研究・開発し、長期的な交流への影響を検証する。

謝辞

本研究を進めるにあたり、ご指導いただいた指導教員ならびに実験に協力いただいた参加者の皆様に感謝する。なお、本研究の一部は JSPS 科研費 (24K03052, 25K00866) の支援を受けた。

参考文献

- [1] Yu, B. and Wright, E.: Paradoxes in intercultural communication, acculturation strategies and adaptation outcomes: international students in Hong Kong, *Journal of Multilingual and Multicultural Development*, Vol. 47, No. 2, pp. 912–929 (2026)
- [2] Abbas, N., Whitfield, J., Atwell, E., Bowman, H., Pickard, T. and Walker, A.: Online chat and chatbots to enhance mature student engagement in higher education, *International Journal of Lifelong Education*, Vol. 41, No. 3, pp. 308–326 (2022)
- [3] Peters, H., Cerf, M. and Matz, S. C.: Large Language Models Can Infer Personality from Free-Form User Interactions, arXiv preprint arXiv:2405.13052 (2024)
- [4] Matšenas, A., Lello, A., Lees, Tõnis, Peep, H. and Tamm, K. L.: Can LLMs Assess Personality? Validating Conversational AI for Trait Profiling, arXiv preprint arXiv:2602.15848 (2026)
- [5] Rammstedt, B. and John, O. P.: Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German, *Journal of Research in Personality*, Vol. 41, No. 1, pp. 203–212 (2007)
- [6] Carciofo, R., Yang, J., Song, N., Du, F. and Zhang, K.: Psychometric Evaluation of Chinese-Language 44-Item and 10-Item Big Five Personality Inventories, Including Correlations with Chronotype, Mindfulness and Mind Wandering, *PLOS ONE*, Vol. 11, No. 2, e0149963 (2016)
- [7] Liu, J.: LlamaIndex, software, available from https://github.com/run-llama/llama_index (2022, accessed 2026-07-03).