LLM搭載家庭用ロボットの対話スタイルがラポールとタスク性能に与える影響

吉田 馨[1,a)]　山本 匠[1,b)]　小橋 洋平[2,c)]　杉浦裕太[1,d)]

概要：
ロボット/AI のコミュニケーションスタイルの変化は，人間とロボット/AI の関係性を変化させ，ロボットの性能（行動計画）やユーザの満足度に影響を与える．しかし，そのスタイル差が実効性能に与える効果は十分に実証されていない．本研究では家庭内ロボットを想定したチャットボットデモアプリケーション CHORD (Collaborative Home-Robot Dialogue) を用いて，LLM 搭載ロボットのコミュニケーションスタイルがユーザとの信頼関係及びタスク遂行に与える影響を明らかにする．具体的には，被験者実験 ($N = 9$) を実施し，タスク志向型・共感型・雑談型の 3 つのコミュニケーションスタイルの変化と，協調タスクにおけるロボット自体のパフォーマンスとユーザの感じ方の関係を調査した．結果，統計的な有意差として，共感型・雑談型はタスク志向型に比べ，ロボットへの「意識の帰属」を強く喚起することが確認された．記述統計においては，共感型が信頼尺度（Trust Scale）で最も高い値（Avg: 2.96）を示し，タスク志向型（Avg: 2.78）と比較してラポール形成に寄与する傾向が見られた．また，雑談型は人間らしさ（Anthropomorphism）において最も高い評価（Avg: 3.16）を得た．

1. はじめに

近年，ロボットの性能は飛躍的に向上しており，社会全体における実用化が加速している．今後 10 年間でロボット産業は年平均成長率（CAGR）が 40 %近くに達する可能性が示唆されており [1]，その活用範囲は製造業や物流などの産業分野にとどまらず，家庭内におけるパーソナルアシスタントとしての潜在市場への進出も急速に進んでいる．このような動向の中で，ロボットが家庭などの私的空間に入る際に，人間とロボットがどのように関係性を築き，協働するのかは，ヒューマン・ロボット・インタラクション（HRI）研究において極めて重要なトピックである．

現在，多くのロボットは「指示＋応答型」の効率重視のインタラクションスタイルを採用しており，最短時間でタスクを遂行することを目的としている．しかし，このようなスタイルでは，ユーザが求める結果とは異なる出力を生じたり，ユーザ側に不要なストレスや違和感を引き起こしたりする恐れがある [2]．今後，ロボットと人との関係がより親密で協調的になるにつれ，単なる効率だけでなく，「人が思う通りに動く」ような体験を重視したインタラクション設計が求められるようになると考えている．

これまでの HRI 研究では，ロボットと人間のラポール（Rapport）関係性や，ロボットの行動がユーザに与える影響に関する検討が行われてきた．本稿においてラポールとは, ロボットと人間との長期的な相互作用を成功させる基盤となる，相互理解や対人的な結びつきを伴った肯定的な関係性を指す [3]．また，近年では大規模言語モデル（LLM）を搭載したロボットとの対話に関する研究も登場している [4]．しかし，それらの多くは会話タスクを想定しており，作業を伴うロボットのタスクパフォーマンスには焦点が当てられていない [5]．さらに，協調的な実働タスクにおける LLM の支援効果をタスク成功率で評価した研究もあるが [6]，コミュニケーションスタイルの違いがパフォーマンスに与える影響については検討されていない．

本研究では，家庭内で人とロボットの協働タスクにおいて，LLM 搭載ロボットのコミュニケーションスタイルがユーザとのラポール及びロボットのタスク遂行に与える影響を明らかにすることを目的とする．具体的には，異なるプロンプト設計によりロボットの対話スタイルをタスク志向型・共感型・雑談型に操作し，それぞれのスタイルによってロボットのパフォーマンス及びユーザの認知負荷，信頼，ロボットの印象がどのように変容するのかを実験的

[1] 慶應義塾大学
[2] 東京大学
[a)] kaoru.yoshida@keio.jp
[b)] imuka06x17@keio.jp
[c)] yohei.kobashi@weblab.t.u-tokyo.ac.jp
[d)] sugiura@keio.jp

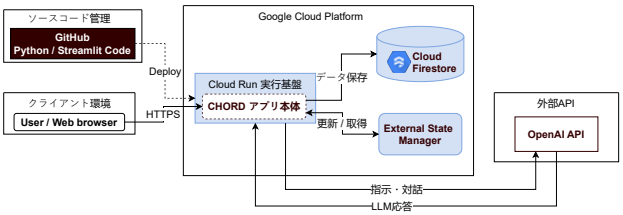

図 1 CHORD のシステム構成

に検証する．

2. 実験システム

本研究では，Streamlit を使用して，家庭内ロボットを想定したチャットボットと対話をするデモアプリケーション「CHORD(Collaborative Home-Robot Dialogue)」を作成し，実験に使用した．CHORD では GPT-4o-mini がロボットの行動計画を生成する．

2.1 ロボットの行動計画とスキルセット

ロボットは，サブタスクの実行を繰り返すことで最終的なタスクの実行を目指す．ロボットは，ユーザの指示の下，実行すべきサブタスクを分解し，定義されたスキルセット（AvailableSkills）から適切なアクションを選択・組み合わせて，FunctionSequence として出力する．ユーザは，これをアプリケーションの「ロボットの行動計画」部分で見ることができる．サブタスクの計画は，プロンプトに定義されたルールに従って自律的に又はユーザの誘導によって構成される[*1]．ロボットが選択可能なスキルは，find，pick up，take など 11 項目に限定した．また，繊細な作業や危険を伴う行為（例：刃物の操作，水を入れるなど）はユーザが代行することを前提とした．

2.2 アプリケーション

アプリケーションでは，トップページ，タスク志向型ページ，共感型ページ，雑談型ページが順番に遷移する．トップページにはアプリの概要と説明動画等を配置し，タスク志向型ページ，共感型ページ，雑談ページは指定されたタスク・タスク完了のイメージ写真・チャット欄・現在の状態タブ・操作パネルで構成した．

タスクは，(a) ごはんのテーブル準備と (b) もらった花束を活けるから自動でランダムに選ばれ，それに対応したタスク完了のイメージ写真が画面上部に表示される．

チャット欄に指示や発言を入力すると（図 2(A)），内容が OpenAI GPT-4o-mini の API に送信され，10〜30 秒ほどで回答が表示される．サブタスクの実行を繰り返すことで最終的なタスクの達成を目指すこととした．会話の中でサブタスクの計画が判断されると，2.1 の AvailableSkills

[*1] プロンプトは https://github.com/kochamii-312/chord.git を参照

図 2 CHORD のページ構成

から成る FunctionSequence が「ロボット行動計画」に表示され，「実行: <次のアクション>」ボタン（図 2(B)）を押すとロボットがアクションを実行し，状態が更新される．サブタスクが数ステップから構成される場合は，ステップ数の回数分段階的にボタンを押す．もし計画が意図と異なる場合は，「実行: <次のアクション>」ボタンを押さず，チャットで修正を指示することができる．

現在の状態タブにはロボットの様子・環境（モノの配置）・タスク目標が表示される（図 2(C)）．この情報は External State Manager (ESM) が操作・管理し，LLM に API で送信される．LLM はこの情報をもとにサブタスクを計画し，ユーザが実行ボタンを押すと ESM を通して更新される．実際にロボットがいないことで状況がわかりにくいのを補填する目的で設置した．

3. ユーザスタディ

3.1 タスクデザイン

被験者は，ロボットと協力して，(a) ごはんのテーブル準備 (b) 花束を花瓶に活ける，の 2 つの家庭内タスクを実施する．両タスクは多段階の手順と判断が必要であり，壊れ物の扱いや順序の判断等の困難な要素を含んでいる．

3.2 コミュニケーションスタイルの分類

以下の 3 つのコミュニケーションスタイルを設定した．

(a) タスク志向型（Task-oriented）は，最小限の情報伝達に徹し，タスクの進行に必要な発話のみを行う．雑談や情緒的反応を避け，効率を重視する．

(b) 共感型（Empathetic）は，共感的な発言（例：「素敵ですね」「疲れていませんか？」）やユーザへの配慮を行う．タスクから逸脱せず励ましや労いの発言を加える．

(c) 雑談型（Small-talk）は，積極的に話題を提示し，環境内のオブジェクトやユーザの行動・感情についてタスクから完全に逸脱する発言，いわゆる雑談を行う．

これらはすべてラポール形成という意図で共通しているが，形成理由が異なる．各スタイルに応じて異なるプロンプトを与えることで，コミュニケーションスタイルを変化させた[*2]．各スタイルに応じたプロンプト内では，発話の

[*2] プロンプトは https://github.com/kochamii-312/chord.git を

トーン，タスク実行のタイミング，質問の形式，発話内容の構造が詳細に制御されている（XML タグによる制御，FunctionSequence の出力条件など）

3.3 仮説

具体的な仮説は以下の 3 つである．

(1) 効率を優先するタスク志向型は，ユーザに相対的に高いストレスを与える一方，共感型・雑談型はタスクに関係のない対話要素により効率は低下するものの，ユーザのストレスを軽減しパフォーマンスを評価を向上させる．
(2) 共感型は精神的なサポートを行いつつタスクに寄り添う姿勢が信頼（Trust）を醸成し，ラポール形成につながる．
(3) 雑談型は，「人間らしい不完全さ」が生命感（Animacy）を喚起し，好感度（Likability）を高めてラポール形成を促進する可能性がある．

3.4 実験手順

実験はタスク志向型条件から開始される．被験者は，5 枚のイメージ写真の中から自身のイメージに最も近い画像を 1 枚選択する．LLM は，これらの画像の情報は持っておらず，あくまでイメージを掴むための参考として利用することとした．次に，「指定されたタスク」に表示された指示文をコピーしてチャット欄に貼り付けて送信することで，会話を開始する．被験者は，質問に答えたり指示を出したりと会話を進め，タスクに関係のない雑談をすることもできる．LLM が行動計画を表示した際は，行動計画の内容を確認し，問題がなければ「実行: <次のアクション>」ボタンを押して段階的にサブタスクを進める．指定されたタスクが完了したと思ったら，操作パネルにある「タスク完了」ボタンを押して会話を終了する．その後，画面下部に評価フォーム（アンケート）が表示され，記入して「評価を保存」ボタンを押すと，「次の実験へ」ボタンが表示される．

3.5 実験設計とデータ収集

実験に同意した参加者は，20〜60 代の男女 9 名であり，任意の場所からアクセス可能な Web アプリ上で実施した．すべての被験者がタスク志向型，共感型，雑談型の条件を実施した．

被験者は実験参加に同意すると，実験アプリ CHORD へのリンクを取得し実験へ進むことができる．記録したデータは被験者氏名，年齢・性別（任意），生成 AI の出力，チャットのやりとり本文，ロボットの状態，環境，作業時間，アンケート回答である．被験者氏名はニックネームでも良いとした．被験者氏名，年齢・性別は実験参加同意フォームで管理し，それ以外はアプリを Cloud Firestore に接続して状態を保存した．

3.6 評価指標

被験者はタスク実施後，以下の評価指標を用いてロボットの印象とパフォーマンスを評価する．

3.6.1 ロボットのパフォーマンス

ロボットのパフォーマンスは，Yang ら [5] の手法に依拠し，タスク所要時間を用いたパフォーマンススコアと，アンケートで示される自己評価の 2 点を用いて測定した．客観指標としては，ユーザのタスク所要時間（秒）を用いてタスク効率を計算する Yang ら [5] の手法を採用した．

$$S_i = \frac{t_{\max} - t_i}{t_{\max} - t_{\min}} \tag{1}$$

主観指標としては，認知負担を計る NASA-TLX の「Performance」項目を採用した．

3.6.2 ユーザの主観的評価

信頼尺度としては，Merritt の Trust Scale を採用した [7]．これは「自律システムに対する感情的信頼感」を測るための心理学的に検証された指標で，HRC 研究でよく使用される．認知的負荷としては，NASA-TLX の全 6 指標を用いた．ユーザビリティの評価を可視化するために採用した．ロボットの印象評価としては，GodSpeed Questionnaire [8] の全 5 指標を用いた．実験参加者は，こうした観点に沿った複数の質問にリッカート尺度（5 段階）で回答した．

評価データ及び実行ログはすべて Cloud Firestore に記録され，プロンプトスタイル別に整理されて保存される[*3]．

4. 結果と考察

各条件の違いを検証するため，3 つのスタイルの結果が欠陥なく得られた $N = 9$ のデータを対象に各質問について反復測定 ANOVA を実施したところ，多くの指標では条件の主効果に有意差が認められなかったが，「意識を持たない-意識を持っている」という項目においてのみ，有意な主効果が確認された（4.4 で詳説）．その他の有意差が認められなかった項目については，統計値として IQR（四分位範囲），Q1（第一四分位数），中央値，Q3（第三四分位数）を含む四分位数，平均値を参照し，その傾向を以下に分析した．なお，各条件の略称を以下で表記し，質問ごとの平均値と標準偏差を示す．タスク志向型：T（Task-oriented），共感型：E（Empathy），雑談型：S（Small-talk）．

4.1 ロボットのパフォーマンス

客観的指標であるパフォーマンススコアの平均値は T: 0.63（std: 0.33），E: 0.60（std: 0.33），S: 0.72（std: 0.29）となり，雑談型が最もタスク効率の良い結果となった．ま

参照

[*3] アンケートの質問項目は https://github.com/kochamii-312/chord.git を参照．

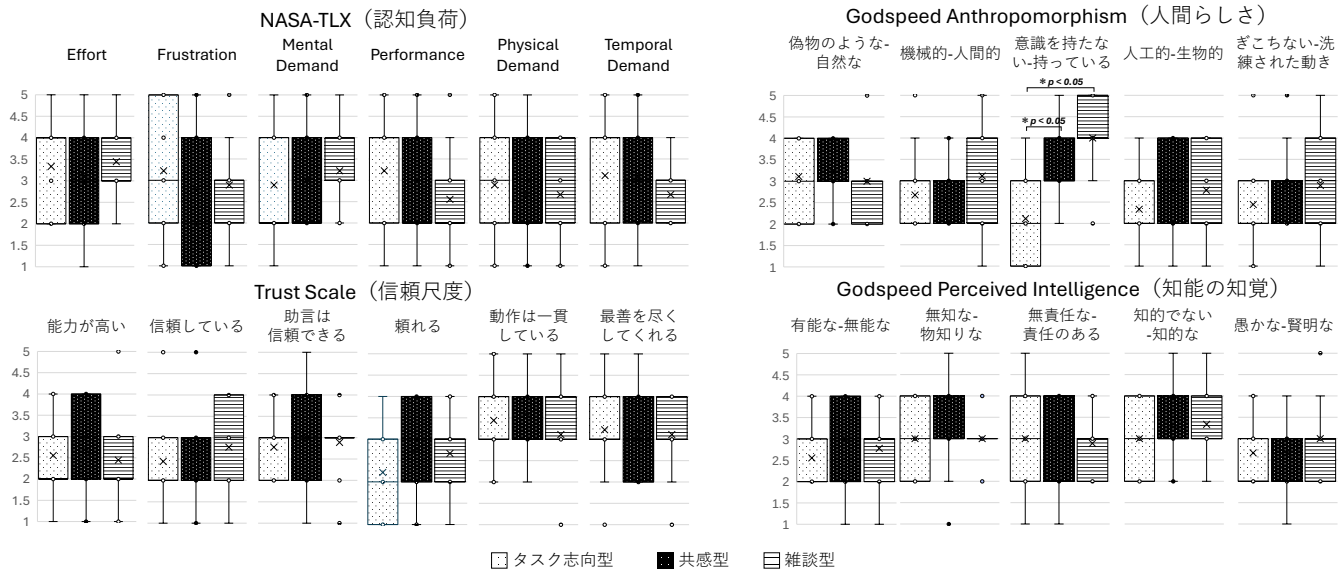

図 3 3 条件における各指標の箱ひげ図（一部指標略称，アンケートの質問項目は `https://github.com/kochamii-312/chord.git` を参照）×: 平均値/-: 中央値

表 1 3 条件（T: タスク志向, E: 共感型, S: 雑談型）における各指標の平均値および標準偏差．括弧内は標準偏差を表す．アンケートの質問項目は `https://github.com/kochamii-312/chord` を参照

	NASA-TLX						Trust Scale						Godspeed Animacy				
	Effort	Frustration	Mental Demand	Performance	Physical Demand	Temporal Demand	能力が高いと信じる	信頼している	助言は信頼できる	頼れる	動作は一貫していると思う	最善を尽くしてくれる	死んでいる-生きている	活気のない-生き生きとした	機械的-有機的	不活発-対話的	無関心な-反応のある
T	3.33(1.05)	3.22(1.55)	2.89(1.29)	3.22(1.40)	2.89(1.45)	3.11(1.52)	2.56(0.96)	2.44(1.07)	2.78(0.79)	2.22(1.03)	3.44(1.07)	3.22(1.13)	3.00(0.67)	3.00(0.94)	2.56(0.83)	3.33(1.15)	3.33(0.82)
E	3.11(1.20)	2.78(1.62)	3.00(1.25)	3.00(1.41)	2.67(1.33)	3.11(1.29)	2.67(1.15)	2.56(1.17)	3.00(1.15)	2.78(1.13)	3.56(0.83)	3.22(1.03)	3.44(0.96)	3.56(0.96)	2.67(0.94)	3.44(1.07)	3.56(1.17)
S	3.44(0.83)	2.89(1.10)	3.22(0.92)	2.56(1.26)	2.67(1.15)	2.67(0.67)	2.44(1.17)	2.78(1.03)	2.89(0.87)	2.67(0.94)	3.11(1.29)	3.11(0.99)	3.22(1.31)	3.33(0.94)	3.00(1.25)	3.33(1.15)	3.44(1.34)

	Godspeed Anthropomorphism					Godspeed Likeability					Godspeed Perceived Intelligence					Godspeed Perceived Safety		
	偽物のような-自然な	機械的-人間的	意識を持たない-持っている	人工的-生命的	ぎこちない-洗練された動き	嫌い-好き	親しみにくい-親しみやすい	不親切な-親切な	不愉快な-愉快な	ひどい-良い	無能な-有能な	無知な-物知りな	無責任な-責任のある	知的でない-知的な	愚かな-賢明な	不安な-落ち着いた	動揺している-冷静な	平穏な-驚いた
T	3.11(0.87)	2.67(1.05)	2.11(0.99)	2.33(0.82)	2.44(1.17)	3.11(0.99)	3.11(0.87)	2.78(1.13)	3.11(0.74)	3.11(0.99	2.57(1.05)	3.00(0.82)	3.00(1.15)	3.00(0.82)	2.67(0.82)	2.89(0.99)	3.44(0.50)	2.89(0.87)
E	3.22(0.79)	2.56(0.83)	3.44(0.83)	2.78(0.79)	2.89(0.99)	2.89(1.20)	3.44(1.17)	2.78(1.15)	2.89(0.99)	3.22(1.03)	3.00(1.05)	3.11(1.10)	3.00(1.15)	3.33(0.94)	2.67(0.94)	3.44(0.68)	3.44(0.68)	2.67(0.47)
S	3.00(1.15)	3.11(1.37)	4.00(0.94)	2.78(1.03)	2.89(1.20)	3.11(0.99)	3.56(0.96)	3.22(1.03)	2.89(0.99)	3.11(0.87)	2.78(0.92)	3.00(0.67)	2.89(0.74)	3.33(0.82)	3.00(0.94)	3.44(0.68)	3.44(0.68)	3.00(0.00)

た，主観的な指標である NASA-TLX の Performance では，平均値は T: 3.22（std: 1.40）が最も高く，E: 3.00（std: 1.41），S: 2.56（std: 1.26）が最も低い値を示した．

Pineda ら [9] によると，雑談のある方がタスク所要時間が長くなる結果を示したが，今回はこれとは異なる結果となった．これは，実験の設計上，雑談型ページが最後の条件であり，被験者が疲労を感じて早く会話を終わらせてしまった可能性がある．

4.2 NASA-TLX（認知負荷）

NASA-TLX の質問全体の平均値は，T: 3.11（std: 1.40），E: 2.94（std: 1.37），S: 2.91（std: 1.06）となり，タスク志向型が最も高い値を示した．Effort と Mental Demand は雑談型が一番高く，他の質問はタスク志向型が最も高かった．とくに Frustration はタスク志向型で Q3 が 5 に達していた．

4.3 Trust Scale（信頼尺度）

Trust Scale の質問全体の平均値は，T: 2.78（std: 1.10），E: 2.96（std: 1.14），S: 2.83（std: 1.08）となり，共感型が最も高い値を示した．共感型では中央値も最も高く，「自分のことを理解してくれるロボットは信頼しやすい」という HRI の文献と一致した．信頼尺度の中でも「動作（ふるまい）の一貫性」は高く，自由記述では「タスク型と優秀さが全く違っていてびっくりした」「性格以外にも違いがあるのか気になった」というコメントもあった．共感的であるとロボットの行動の質も高く感じられる「ハロー効果」とも一致する [10].

4.4 Godspeed Questionaire（ロボットの印象）

Animacy（生命感）の質問全体の平均値は T: 3.04（std: 0.94），E: 3.33（std: 1.07），S: 3.27（std: 1.22）となり，共感型が最も高い値を示した．タスク志向型はすべての質問で他のスタイルより平均値が最も低く，指示中心で反応が少ないため，生命感のない印象を与えたと考えられる．

雑談型では「機械的な-有機的な」のみ平均値が雑談型が一番高く，他は共感型が一番高かった．

Anthropomorphism（人間らしさ）の質問全体の平均値は，T: 2.53 (std: 1.05), E: 2.98 (std: 0.91), S: 3.16 (std: 1.23) となり，雑談型が最も高い値を示した．特に「意識を持たない-意識を持っている」という項目においては，有意な主効果が確認された（$F(2, 16) = 10.9701$, $p = .0010$）．この項目について Bonferroni 補正付きの事後比較を行ったところ，共感型はタスク志向型より優位に高い意識性の評価を示した（$t(8) = 4.62$，$p = .0051$，$Hedges'sg = 1.31$）．同様に，雑談型もタスク志向型より優位に高かった（$t(8) = -3.90$, $p = .0136$, $Hedges'sg = 1.75$）．一方で，共感型と雑談型の間には有意差は見られなかった（$t(8) = -1.25$, $p = .7399$）．これらの結果から，タスクに特化した機械的な応答に比べ，共感的な応答や雑談を含む応答は，ユーザにロボットの「意図性」や「心的状態」を強く帰属させることが示唆される．雑談によってロボットがタスク中心の応答から離れて話題を広げるため，ユーザはロボットを「意図を持つ主体」に近いものとして知覚したと考えられる．共感型も人間らしさを高めるが，雑談型のような自律性や脱線行動が少ないため，その上昇は限定的であった．

Likeability（好感度）の質問全体の平均値は，T: 3.04（std: 0.97），E: 3.09（std: 1.13），S: 3.18（std: 1.00）となり，雑談型が比較的高い評価となったが，特に有意差は認められなかった．

Perceived Intelligence（知能の知覚）の質問全体の平均値は，T: 2.84（std: 0.92），E: 3.02（std: 1.04），S: 3.00（std: 0.84）となり，共感型が最も高い値を示した．タスク志向型は，他のスタイルと違って中央値・平均値ともに 3 を超えず，予想に反して低く抑えられた．共感型では 5 つの質問のうち 4 項目で Q3 が 4 に達していた．これは，共感的だとロボットの行動の質も高く感じられる「ハロー効果」[10] が発揮された可能性がある．

Perceived Safety（安全性の知覚）は，ロボットのふるまいではなくユーザ自身がどのような心の状態になったかを評価した項目である．Perceived Safety の質問全体の平均値は T: 3.07（std: 0.86），E: 3.19（std: 0.72），S: 3.30（std: 0.60）となり，雑談型が最も高い値を示した．

4.5 自由記述（定性評価）

タスク志向型では，「ないだけですませるのは低レベル」「代替案や申し訳なさが欲しい」といった対話の柔軟性の低さや，「提案をしてくれると，それが的を得ていなくても，信頼関係の醸成につながるかもしれない」といった，積極的な提案姿勢を求める意見が得られた．

共感型では，情緒的なつながりを評価する肯定的な意見が多く見られた．「思いが通じ合えるようになった」「自然なやりとりで気持ちよくタスクできた」「人間の話を認めてくれるところがすごくいい」「向学心と親しみやすさでその場の情報・状況を学んで成長」「こんな AI と暮らしたい」といった声が挙がった．「受け身ではなく，会話の中で自然に協力できるような空気を作ってくれた」といったロボットの印象も見られた．一方で，過度な配慮に対し「子ども扱いされているように感じた」「自立支援されているようだった」と自立性を損なう印象を持った参加者もいた．

雑談型では，「初めて人間らしさを感じた」「雑談のレベルが高く自然」「信頼関係を築くためには雑談力が必要」など，人間味や信頼構築の観点で高い評価が得られた．また，「会話が離れていっても戻してくれる」点に安心感を覚える意見もあった．一方で，タスク遂行を重視する場面では「余計な会話が多くて疲れた」「指示以外の回答もしないといけないのは徒労」といった批判も見られた．これに対し，「シンプルな事をお願いしたい場合は，タスク志向型が良いが，他のケース（パーティー等）の場合であれば，雑談型」という意見もあり，状況に応じた提案がユーザに好印象を与えることが示唆された．

全ての条件を通じて，システムやロボットの機能的制約に対する言及が多数確認された．「ロボットがいないので状態がわからない」，段階的実行に対し「動作を確認するのが苦痛」という意見があった．また，把持能力等の制約に対し，ルールによる強制が「無能感」や「いらいら」を増幅させたことが分かった．一方で，「できないことはできないと言ってくれてありがたい」という受容的な意見や，制約を「無能さ」ではなく「作法」や「おもてなし」と捉える意見，さらには「ロボを幼い子として捉え，助けてあげれば楽しく作業ができる」といった，制約を肯定的に再解釈（リフレーミング）する様子も観察された．

自由記述の結果から，共感型ではユーザの話を認め，協力して学ぼうとする姿勢が評価された．雑談型では，中立的な AI らしさを脱却し，人間味を与えることができた．また，雑談型では仮説のような「人間らしい不完全さ」よりむしろ「アイデア性」が評価されたことが推測できる．

また，ユーザが期待する動作水準と実際のロボットの能力（無能さ）とのギャップが，苛立ちなどの負の感情を生む要因であることが確認された．しかし，ロボットの未熟さが「助けてあげたい」という感情を生み，ギャップを補完しうることが示唆された．これは期待度と真の能力の大小関係がインタラクションを継続するかに影響する「適応ギャップ仮説」によっても説明できる [11] [12]．幼い口調のコミュニケーションスタイルや外見やしぐさ等の非言語情報を通じて意図的に期待値を下げ，「助けてあげたくなる」印象を付与することがユーザの受容性を高め，機能的不全をカバーする有効なアプローチになると考える．加えて，自由記述のネガティブなコメントはデモアプリケーションの不具合によるものが多かったが，同様に家庭環境において，ロボットが物理的にスムーズな動作を保証する

ことは現状困難であり，物理的動作の不具合による負の感情が生まれることは十分想定される．したがって，不完全さを前提とした HRI 設計が重要となると考える．

5. さいごに

本研究では，家庭内ロボットにおける LLM の対話スタイルが，ユーザとのラポール形成及びタスク遂行に与える影響を検証した．以下に 3.3 の仮説に対する結論を示す．

(1) タスク志向型は NASA-TLX の総合的な認知負荷平均値が 3 条件中で最も高く（3.11），ストレスが高いという仮説は支持された．また，Performance は最も高い評価（3.22）を得た．効率を最優先し情緒的サポートを排除したスタイルがユーザに心理的な負荷をかける反面，タスク遂行に対する確実な手応えを与えたことを示唆している．

(2) 共感型は Trust Scale において 3 条件中で最も高い値（2.96）を示し，Perceived Safety においても高い評価を得ている．特筆すべき点として，Perceived Intelligence においても最も高い評価（3.02）が得られた．これは，ロボットが情緒的に寄り添う姿勢を見せることで，ユーザがその能力や知能までも高く評価する「ハロー効果」が生じたと考えられる．統計的にも，「意識を持っている」において，タスク志向型に対し有意に高い評価（$p < .01$）を得ている．これらの結果は，共感的な対話が信頼関係とラポール形成に有効であることを示している．

(3) 雑談型は，Anthropomorphism 及び Likeability において高い評価を得ており，「人間らしい不完全さが好感を生む」という仮説の前提部分は支持された．自由記述においても「人間味を感じた」「提案性が良かった」といった肯定的な意見が見られた．しかし，NASA-TLX の Effort が高く，Frustration が最も広く分布するなど，タスク遂行においては雑談が「阻害要因」と捉えられる側面も確認された．雑談型は人間らしさや親しみやすさを醸成する一方で，タスクの文脈によっては必ずしも円滑な協調（ラポール）に直結しない可能性が示唆された．

本実験では共感型が総合的に優れていたが、長期的な失敗の許容や飽きの回避には限界があると考えられる。継続的な利用には、対話自体の楽しさが本質的に重要であり、雑談要素の導入が不可欠である。例えば、ユーザの個性や環境情報を反映した文脈的な雑談を行うことで、関心の高い有益な情報提供が可能となり、より強固なラポール形成と利用継続につながると期待される。

HRI 設計への示唆として，現状のロボットの物理的制約（不完全さ）を補完するためには，「指示+応答型」のコミュニケーションスタイルだけでは不十分である可能性がある．むしろ，共感型のように「親しみを持ってユーザと協力して学習する」ことや，雑談型のように「アイデアを提案する」ことが機能的不全に対する受容性を高める鍵になると考えられる．

なお，本実験ではスタイルの提示順序が固定されており，コミュニケーションスタイルの名称によるバイアスの影響も否定できない．今後は，提示順序のランダム化を含めた厳密な検証を行うとともに，長期的なインタラクションを通じたラポール変容の解明が求められる．

謝辞
調査には，東京大学松尾・岩澤研究室 LLM コミュニティのプログラム LLMATCH にご協力頂きました．

参考文献

[1] Ding, R.: Application and optimisation of intelligent control system for home robots, *Applied and Computational Engineering*, Vol. 102, pp. 102–107 (2024).
[2] Reeves, B. and Nass, C.: *The media equation: How people treat computers, television, and new media like real people*, Cambridge University Press, Cambridge, UK (1996).
[3] Lin, T.-H. et al.: Connection-Coordination Rapport (CCR) Scale: A Dual-Factor Scale to Measure Human-Robot Rapport, *2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, IEEE (2025).
[4] Kim, C. Y., Lee, C. P. and Mutlu, B.: Understanding large-language model (LLM)-powered human-robot interaction, *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction* (2024).
[5] Ye, Y., You, H. and Du, J.: Improved trust in human-robot collaboration with ChatGPT, *IEEE Access*, Vol. 11, pp. 55748–55754 (2023).
[6] Rahman, S. M.: Trust-Based Modular Cyber–Physical–Human Robotic System for Collaborative Manufacturing: Modulating Communications, *Machines*, Vol. 13, No. 8, p. 731 (2025).
[7] Merritt, S. M.: Affective processes in human–automation interactions, *Human Factors*, Vol. 53, No. 4, pp. 356–370 (2011).
[8] Bartneck, C. et al.: Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots, *International Journal of Social Robotics*, Vol. 1, No. 1, pp. 71–81 (2009).
[9] Pineda, K. T., Brown, E. and Huang, C.-M.: "See You Later, Alligator": Impacts of Robot Small Talk on Task, Rapport, and Interaction Dynamics in Human-Robot Collaboration, *2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, IEEE (2025).
[10] Thorndike, E. L.: A constant error in psychological ratings, *Journal of Applied Psychology*, Vol. 4, No. 1, pp. 25–29 (1920).
[11] Aronson, E. and Linder, D.: Gain and loss of esteem as determinants of interpersonal attractiveness, *Journal of Experimental Social Psychology*, Vol. 1, No. 2, pp. 156–171 (1965).
[12] Komatsu, T., Kurosawa, R. and Yamada, S.: How does the difference between users' expectations and perceptions about a robotic agent affect their behavior? An adaptation gap concept for determining whether interactions between users and agents are going well or not, *International Journal of Social Robotics*, Vol. 4, No. 2, pp. 109–116 (2012).