

[研究助成 (A)]

RLHF を用いた自閉症児の個性に合わせた LLM の最適化と
寄り添う対話ロボットへの実装

2241034

研究代表者
(助成金受領者)

共同研究者

金沢大学 人間社会研究域

追手門学院大学 理工学部

金沢工業大学 工学部

教授

准教授

教授

吉 村 優 子

高 橋 英 之

出 村 公 成

[研究の目的]

本研究の目的は、自閉スペクトラム症（ASD）児の社会生活の質を高めるため、子どもの自己肯定感を維持・向上させる声かけを持続的に行う対話ロボットを開発・検証することである。ASD はコミュニケーションの困難を主訴とする神経発達症であり、物事の感じ方や情報処理の仕方が定型発達と異なることから、自己肯定感の低下や社会不適応を生じやすい。本研究では、当事者を単なる「被験者」ではなく開発の主體的参加者と位置づける参加型アクションリサーチ（CBPR）の枠組みを採用し、(1) 子どもの特性に合った発話の生成、(2) 適切なタイミングでの発話の表出、という二つの技術的課題の解決を目指した。(1) については RLHF（人間のフィードバックによる強化学習）を用いた大規模言語モデル（LLM）の個別最適化を、(2) については対話ロボットを集団共有アバター化し、子どもや養育者による半自律制御を行う枠組みの構築を計画した。これらを通じて、人間と機械の調和した関係構築に資する新しいコミュニケーション支援ツールの可能性を示すことを長期的なビジョンとして設定した。

[研究の内容、成果]

1. ワークショップ・インタビューによる理想
発話例の収集

ASD 児を含む発達障害のある子どもを対象に、ロボットによる理想的な応答（声かけ）に関するインタビューおよびワークショップを実施した。本研究は当事者を開発の主體的参加者と位置づける参加型の研究実践を重視しており、その一環として対話的フィードバックを得る形で研究基盤を構築した。実際の教育・療育フィールドとの関係づけを意識し、特異な興味・才能をもつギフテッドの子どもを対象としたワークショップや、多様な特性の子どもが集まる放課後等デイサービスでのロボットイベントを開催し、子ども自身が大規模言語モデルのプロンプトを書き換えて「自分好み」のロボットをつくり、それを評価・分析する“研究員”的な参加の場を設計した。また支援者向けのオンラインサイエンスカフェを通じて成果報告と建設的議論を重ねた。一方で、対象児の特性や保護者の配慮の必要性から研究協力への同意取得が難しい事例も少なくなく、継続的な信頼関係の構築と研究手法の工夫が今後の課題として明確になった。

2. ロボットによる「ほめ」の種類が動機づけ
とロボット印象に与える影響の調査

発達支援において「ほめ方」は行動的介入の

コアエレメントとして位置づけられており、具体的な行動を観察・記述する「説明的ほめ」と全般的な肯定を行う「一般的ほめ」が代表的手法として知られている。先行研究では ASD 傾向の子どもに対して「説明的ほめ」が学習を促進する可能性が示唆されているが (Polick et al., 2012), ロボットが行う「ほめ」の効果については未検討であった。そこで、本研究では、どちらのほめ方が子どもの動機づけとロボットへの印象に効果的かを検討した。

小学校5年生から高校1年生8名(女児1名・男児7名, うち ASD 診断あり2名, ADHD・不安症1名, 社交性不安症1名)を対象に, 2体のオリジナルロボット(「太陽ロボット」「北風ロボット」と命名, 図1)による2種類の「ほめ」動画を視聴してもらった。場面は①分数の宿題に取り組んでいるとき, ②5キロのマラソンを走っているときの2場面を設定した。一般的ほめ(例:「分数の勉強頑張ってるね, すごいよ。きみなら最後までできる!」)と説明的ほめ(例:「分数の勉強をしているんだね, 6ページのうち3ページ, すごく難しかったけど, きみはとけたんだ!」)を比較した。評価は Google フォームを用いた7件法のアンケートにより, 活動への動機づけ項目6項目(「自分には能力があると思った」「もっと努力したいと思った」「やればできるといった」等)とロボットへの印象項目2項目(「ロボットの言葉を嬉しいと感じた」「ロボットが好きだと感じた」)の計8項目で測定した。



図1 調査に用いたぬいぐるみロボット

全体的な傾向として、動機づけは「一般的ほめ」によって高まる傾向がみられた(図2)。一方、ロボットへの印象については2種類のほめ方による大きな差は認められなかった。場面別にみると、宿題場面では、ASDの診断がない子どもたちにおいて、一般的ほめによる動機づけが高く、診断がある子どもたちでは全体的に得点が低い傾向を示した。②マラソン場面では診断の有無と応答種類による交互作用様のパターンがみられた。参加者からは「さりげなく助けてくれる感じがいい」「自分だけに当てはまることを言われると、ちゃんと自分のことを見て励ましてくれていると感じる」「宿題しているときは何も言わないで見守ってくれるだけでもいい」など、声かけの内容・タイミング・関係性に関する多様な視点が示された。上記の調査からは、ASD傾向の子どもが「説明的ほめ」を一律に好むわけではなく、特性・場面・年齢によって有効なほめ方が異なる可能性が示唆された。

3. ロボットによる「応援」の種類と個人特性の関係に関するウェブ調査

RLHFの個別最適化に向けた基礎知見として、ロボットによる「応援」の内容とユーザーのパーソナリティ特性との関係を検討した。ウェブアンケート調査(n = 32, 平均年齢36.2歳, 範囲10~59歳)において、ロボットによる4種類の「応援」動画(V1「成功のイメージを頭に鮮明に浮かべたら、必ず成功をおさめ

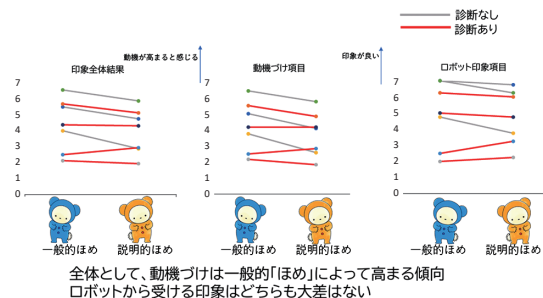


図2 ほめ方の違いと動機付け、ロボットへの印象の関係

るよ!」, V2「頑張れ, きみならできる!」, V3「どんな結果になっても僕は君の味方だよ」, V4「どんな結果になっても君なら大丈夫」を呈示した。印象の測定には, 親密さの評定尺度 (小山ら, 2017) から5項目と独自の追加5項目 (「励まされた」「愛情を感じた」「受け入れてもらえた」「プレッシャーを感じた」「悪意を感じた」) の計10項目を8件法リッカート尺度で測定した。パーソナリティ特性は日本語版 Ten Item Personality Inventory (TIPI-J) を用いて評定した。因子分析により低社会的結合項目・高社会的結合項目・負担項目の3因子を抽出し, 各応答との関係を Spearman の順位相関により検討した。

V4「どんな結果になっても君なら大丈夫」は, 友情・励ましの印象 (低社会的結合項目) が最も高く, 負担感が最も低い応答として評価された ($p < .01$)。V3「どんな結果になっても僕は君の味方だよ」は「家に招きたい」「一緒に生活したい」といった高社会的結合項目で最も高く評価された ($p < .01$)。一方, V1「成功のイメージ…」は親しみを感じにくく負担感が最も大きい応答として評価された ($p < .01$)。パーソナリティとの関連については, V1に対して外向性の高い人は低社会的結合得点と正の相関 ($p = 0.357, p = .045$) を示した一方, 神経症傾向の高い人は低社会的結合得点 ($p = -0.371, p = .036$) および高社会的結合得点 ($p = -0.419, p = .017$) と負の相関を示した。不安傾向を伴いやすい対象者には, 「どんな結果でも大丈夫」

という結果への寛容を示す声かけ (V4) が, プレッシャーを与えにくく有効である可能性が示唆された。パーソナリティ特性によって同じ「応援」が全く異なる印象をもたらすという本知見は, RLHF による声かけの個別最適化の必要性を裏付けるものである。

4. 「寄り添い」応答に対する印象の発達差に関するウェブ調査

RLHF の教師信号となるフィードバックデータの収集モデルを確立するための基盤的知見を得る目的で, ロボットの「寄り添い方」を心理カウンセリングの代表的手法に基づき五種類 (「弱さを共有」「無関係な話」「問題解決」「共感」「称賛」) に分類し, それぞれに対する印象を成人 (18 ~ 40 歳) と子ども (10 ~ 15 歳) で比較するオンライン調査を行った。独自に開発したスイッチロボット (大道・高橋ら, 2024) が, 「連立方程式が解けず試験に失敗した」という人間の発話に対して五種類の応答を行う動画を呈示し, その印象を回答してもらった。各応答文は OpenAI 社の GPT-4o のプロンプト調整により作成した。印象の測定には Technology Acceptance Model 尺度 (Heerink et al., 2010) から8項目を抽出し, 独自の2項目 (「ずっと一緒にいたい」「ロボットは信頼できる」) を加えた計10項目を5件法のリッカート尺度で用いた。動画内容の理解を確認するテスト設問に正答した成人112名・子ども110名を解析対象とした。本調査は金沢大学の研究倫理委員会の審査・承認を得て実施した。

主成分分析により10項目をポジティブな印象とネガティブな印象の二因子に圧縮し, 参加者属性 (成人・子ども) とロボットの応答種類を要因とする二要因分散分析を行った。その結果, ポジティブな印象においてのみ交互作用が有意であった ($F(4, 840) = 4.51, p = .005$)。下位検定の結果, 「弱さを共有」「問題解決」「共感」の応答において, 子どもは成人よりも有意にポジティブな印象を抱く傾向が認められた (図4)。

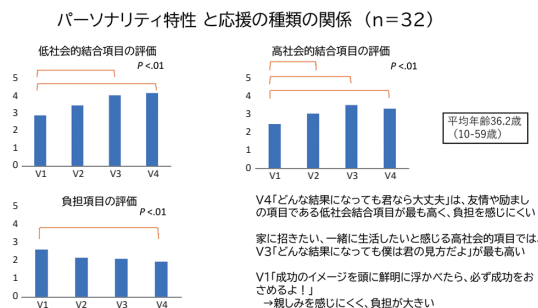


図3 ほめ方の違いと動機付け, ロボットへの印象の関係

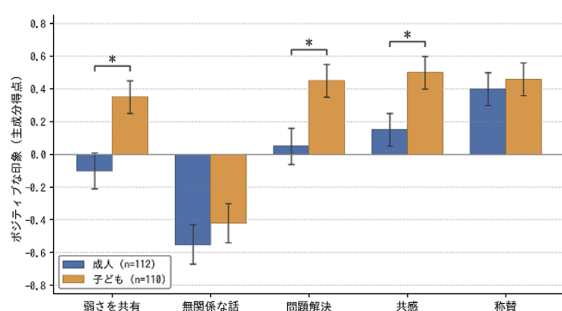


図4 応答に応じたロボットへのポジティブな印象の度合（成人と子どもの比較）。エラーバーは標準誤差、*は下位検定で成人と子どもの間に有意差があった条件を示す。

成人は「問題解決」的な応答に対して子どもほど高い印象を示さない一方、「称賛」的な応答には子どもと同程度のポジティブな印象を示した。この結果は、成人がロボットに「自らの肯定者」としての役割を期待する傾向を示唆しており、ライフステージに応じた承認欲求の変化を反映している可能性がある。

本調査により、課題（1）子どもの特性に合った発話の生成に向けて、子どもがどのような「寄り添い方」に肯定的な印象を持つかという基礎データを取得し、自己肯定感を支える発話タイプの選定に関する知見を得た。課題（2）適切なタイミングでの発話の表出については、人間のネガティブな発話に対する応答の「内容」を中心に検討し、状況に応じた発話内容とタイミングの関係についてさらなる整理が必要であることを明らかにした。これらの知見をもとに、子どもの特性に応じて「いつ・どのような言葉をかけるか」という応答生成条件とタイミング制御の設計指針を整理し、半自律制御による集団共有アバター型プロトタイプの設計と小規模実装、ならびに教育・療育の現場における予備的なフィールド検証へと展開した。RLHFデータの拡充とプロトタイプの個別最適化を進めることで、ASD児の自己肯定感を継続的に支える対話ロボットの実現可能性を示した。

5. 研究全体の到達点と今後の展開

本研究では、ASD児の自己肯定感を支える声かけを行う対話ロボットの開発に向けて、当事者参加型のワークショップ・インタビュー、ロボットによる「ほめ」の種類の比較、「応援」の種類と個人特性の関係、「寄り添い」応答の発達差という多角的なアプローチから基礎的知見を積み上げた。

技術的課題（1）「子どもの特性に合った発話の生成」については、以下の知見が得られた。第一に、「弱さを共有」「問題解決」「共感」といった寄り添い応答に対して子どもは成人より有意にポジティブな印象を持つ一方、「称賛」については成人と差がないという発達段階による差異が明らかになった。第二に、ロボットによるほめの効果は特性・場面・年齢によって異なり、ASD傾向の子どもが「説明的ほめ」を一律に好むわけではないことが示された。第三に、「どんな結果でも大丈夫」という結果への寛容を示す声かけが、不安傾向を伴いやすい対象者に対しては、特に有効である可能性が示唆された。これらをRLHFの教師信号として活用することで、子どもの特性・状況・感情状態に応じた応答生成の個別最適化が可能となる基盤が整った。

技術的課題（2）「適切なタイミングでの発話の表出」については、ロボットがネガティブな発話に対してどのような内容の応答を返すかという「内容」面の検討を中心に進めた。「さりげなく助けてくれる感じがいい」「何も言わないで見守ってくれるだけでもいい」といった子どもからの自由記述は、発話の内容だけでなくタイミングや関係性の質が自己肯定感の維持に深く関わることを示唆しており、「いつ・どのような言葉をかけるか」という応答タイミング制御の設計は次年度以降の重点課題として引き続き取り組む。

研究推進上の課題としては、対象児の特性や保護者の配慮の必要性から研究協力への同意取得が難しい事例が少なくなかった。当事者を「被

験者」ではなく開発の主体的参加者と位置づける CBPR の枠組みのもと、教育・療育現場との継続的な信頼関係の構築と参加しやすい研究手法の工夫が不可欠である。

今後は、上記の知見をもとに、RLHF データの拡充と LLM の個別最適化を進めながら、集団共有アバター型対話ロボットのプロトタイプを教育・療育の現場で本格的にフィールド検証することを目指す。子ども・養育者・支援者が協働してロボットの応答をチューニングしていく半自律制御の枠組みを通じて、ASD 児一人ひとりの個性に寄り添い、自己肯定感を継続的に支える対話ロボットの実現可能性を示していく。

[成果の発表、論文など]

吉村優子・高橋英之：多様な子どもの主体性を引き出すな
んもしないロボットの可能性. 日本ロボット学会誌, 43.
2025 年.

高橋英之・吉村優子：成人と子どもでロボットに期待する
「寄り添い」の違い, 日本心理学会大会, ポスター発表,
2025 年 (国内発表).

吉村優子・高橋英之：多様な特性のある子どもにおける心
地よくロボットと共生する空間に関する研究, 日本認知
行動療法学会 シンポジウム「臨床家がロボット, アバ
ター, 及びビッグデータと連携する方法」, 2023 年 10
月 (国内発表).

(関連発表) 大道・高橋ほか：スイッチロボットを用いた家
電連動型コミュニケーションに関する研究, International
Conference on Social Robotics (ICSR), 2025 年.