

# 物理的 UI を用いた一般ユーザーの LLM 活用促進手法の提案

福田 大翔\* 川辺 航\* 菅野 裕介\*

**概要.** この論文では、一般ユーザーが大規模言語モデル (LLM) を利用するうえで存在する課題を探るとともにその課題を解決できるようなアプローチを検討する。LLM は本来適切な前提を提供することで常識的な振る舞い以外にもユーザー独自の振る舞いをさせることができる。しかし、多くのユーザーは LLM の多機能性やカスタマイズ可能性を知らないために、LLM を活用した課題解決を行うことを困難にしている可能性がある。この問題を解決するため、本研究では物理的な UI を活用した、LLM を利用しつつも全ての入出力をデバイスを通じて行う手法を提案する。また提案デバイスが LLM の柔軟性の理解を促進するかを検証した。ユーザー実験ではデバイスを利用することを通じて LLM に対する認識がどう変わっていったかをアンケートやユーザーの振る舞いから調査した。結果として、このデバイスは一般ユーザーが LLM の可能性に対する理解を深められうることが確認された。

## 1 はじめに

近年、LLM は高性能化しており、様々な用途で利用することが可能になってきている。チャットボットとしてや検索の補助としてだけではなくプログラミングやデータ分析など幅広い用途での使用が可能である [2]。そのため、LLM を十分に扱うことができればユーザー自身の可能性を大きく高めることができる。しかしながら、LLM に対して専門的な知識を持たない一般ユーザーが LLM を活用することは現時点では容易ではない。

本研究では上記課題に対して、一般ユーザーが LLM のカスタマイズ性、多機能性を十分に認識していないことが問題の根本にあると仮定した。LLM は課題に合った前提知識や自分オリジナルの設定、求める出力形式などの周辺情報を提供することで状況にあった知識を活用させた返答を行わせることも可能である。しかし一般ユーザーにとっては、調べ物のような LLM が既に持っている知識を引き出すような使い方が一般的であることが予想される。さらに、一般のユーザーが使いやすい形で提供されている LLM を利用したサービスは、GUI で会話型のインタフェースを採用している場合がほとんどである。一般ユーザーが会話を通して LLM とインタラクションを取る中ではそのようなユーザー自身が LLM をコントロールできることに気づきづらい可能性がある。LLM が持つ本来の可能性を最大限活用するには、システムプロンプトの細かいチューニングや、更には API やオープンソースモデルを利用したプログラミングが必要になる。これらを利用するには当然ながら専門的な知識が必要であり、一般

ユーザーが利用する選択肢としては挙がりづらい。

こうした背景を踏まえると、一般ユーザーであっても手軽に LLM の用途の幅やコントロールできることを知るための手段が求められる。この課題に対する一つのアプローチとして、専門知識を持たないユーザーでも LLM のカスタマイズ性を気軽に体験できるような、物理的・身体的な UI を提供することが考えられる。GUI 上の会話や複雑なプログラミング体験に依存するシステムとは異なり、LLM の用途の幅広さやカスタマイズ性を物理的に体験することでより直感的に LLM の可能性を理解しやすくなることが期待される。先行研究では機械学習の教育目的で物理的な UI の有効性が確認された例がある [4, 17, 5]。しかし、これらは機械学習モデルの一つである分類モデルに対する例であり、LLM のような入出力が複雑なモデルに対する適用は未知数である。本研究では、技術的な専門知識を持たないユーザーが LLM の可能性をどう認識しているかと、物理的な UI を利用したデバイスによりその認識が変化するかについて検証・調査を行うことを目的とする。本研究が提案する物理的 UI・デバイスは図 1 のような外観である。ユーザーはデバイスに音声で指示を与えることでデバイスの操作を行えるほか、デバイスに指示とそれに対応する動きを学習させることができる。言葉と動きのマッピングには LLM を活用しており、指示は具体的な動作を直接表すものだけではなく様々な内容で自由に行うことができる。すなわち、学習させる言葉と動きの対応関係は意味通りである必要がなく、ユーザーが自由に設計できるようになっている。入出力は全てデバイス内で完結されており、PC 操作に不慣れたユーザーでも手軽に扱えるように設計されている。通常モードでは定期的に録音された言葉に対する動作の推論を行い、デバイス自身が実際にその動作を実行する。学

Copyright is held by the author(s). This paper is non-refereed and non-archival. Hence it may later appear in any journals, conferences, symposia, etc.

\* 東京大学



図 1. 提案デバイスの外観.

習モードではユーザーが学習させたい言葉を音声で与え、その後実際にマット上でデバイスを手動で動かすことでその軌跡を言葉に対応する動作として指定する。言葉と動作のペアを例として LLM 指示に与えることで、教えられた言葉そのものだけではなく、それ以外の言葉もそこから類推した様々な動作を行うようになる。本論文ではこのデバイスを用いたユーザー実験を行い、一般ユーザーの LLM への認識の現状とデバイス利用後の認識の変化について調査を行った結果について報告する。実験参加者は利用方法の説明の他「デバイスを活用した遊びを作る」という指示のみを与えられ、デバイスを利用した開発を行った。実験前には LLM に対する認識の現状を調査するアンケートを行い、実験後には同じアンケートのほか、参加者数名に対してデバイス全般や LLM への認識に関するインタビューを行った。

## 2 関連研究

### 2.1 物理的な UI の ML への活用

物理的な UI を活用し、一般ユーザーへの機械学習の理解を促進する手法については様々な試みが行われている。特に機械学習の学習を目的とした手法が多く存在する [4, 5, 17]。絵をバイナリ分類するモデルをデバイスを通じて作成する中で分類モデルを利用するのに必要な過程を理解でき、ユーザー自身が考えを元にモデルの改善をできるもの [5]、スポーツにおいて自身の運動データを記録しそれを元に動作の成否を判定するモデルをつくることを通じてユーザー自身の特性を活用したモデルを作る体験ができるもの [17]、ジェスチャーを分類するモデルを作成することを通じてモデルへの理解を深められるもの [4] である。また、エッジデバイス上でモデルの処理を行うデバイスを用いてモデルを作成することでユーザビリティやモデル作成の効率を検証したものが存在する [10]。

これらの研究は物理的な UI を活用することで、

より容易に機械学習を扱えることを目指したものである。特に教育にまつわる研究の結果では、機械学習にまつわる物事を知識のない一般ユーザーに学習させる点において物理的な UI は一定の効果があるとされている。一方、これらの研究はどれも分類モデルを通じた学習、改善を試みたものである。LLM は入出力可能な内容が複雑であり、LLM でも物理的な UI が有効かは未知数である。本研究では機械学習モデルとして LLM を用いる点が既存手法と大きく異なる。UI の改善や直観的な扱いやすさはもちろん、インタラクション手法の幅の広さを生かして広げ身体的な入力を行えるようにし、またそのような入出力とテキストとの間の変換を適切に行うことで、LLM においても物理的な UI の有効活用を目指していく。

### 2.2 プロンプトエンジニアリング

プロンプトエンジニアリングは LLM に与える命令を最適なものにする課程やその技術のことであり、LLM の動作をよりユーザーの望むものに近づけることができる。これには様々な手法が存在する。まず与えるテキストそのものを調整する手法がある [1, 11, 14, 13, 6]。また、ツールを利用する手法も存在する。一つは命令の生成にも LLM を利用する手法 [16, 15]、もう一つは LLM ではない小型のモデルを利用する方法 [12, 9, 3] である。その他、LLM に入力された命令文のベクトル量にアクセスできる場合には、それを活用する手法も存在している [7, 8]。

本デバイスではプロンプトエンジニアリングそのものはユーザーから見えない形にする一方でそこから実現される LLM の柔軟性をまず知ってもらうことを目指す。内部の処理で LLM にタスクに関する例を与えることを活用しつつ、利用者が入出力からそのカスタマイズ性を実感できるようにする。

## 3 提案手法

我々のデバイスは音声で指示を与えることで言葉の内容に基づいてデバイスの操作を行うことができるほか、デバイスに言葉とそれに対応する動きを学習させることができる。これは、一般ユーザーに LLM のカスタマイズ性や多機能性を実感してもらうことを意図したものである。カスタマイズ性については、ユーザーがモデルに独自の内容を学習させ、それがデバイスの挙動として反映されるという過程で感じることができる。LLM は、入力された言葉の内容によらず何かしらの推論結果を出力する。そのため学習させる言葉の意味と動きが対応する必要はなく、両者の間に明確な関連性がなくても問題ない（「アイスクリーム」と言われたら右に進む、など）。ユーザーが自由に学習内容を設計する過程でカスタマイズ性を感じることが我々の狙いである。多機能

## 物理的 UI を用いた一般ユーザーの LLM 活用促進手法の提案

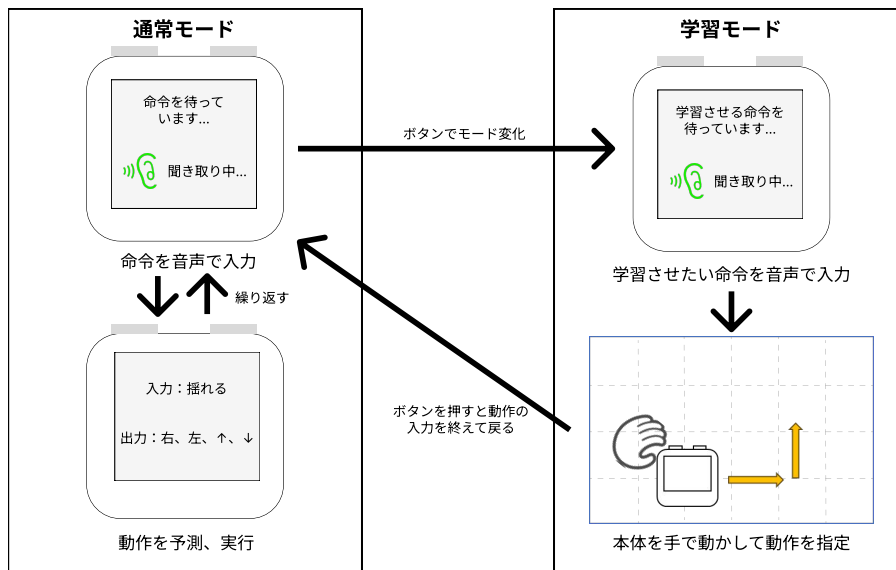


図 2. デバイスの動作フロー。通常モードでは音声の命令に対して動作の推論を行い，学習モードでは命令とそれに対する動作を学習させることができる。

性については、現在比較的普及している LLM アプリケーションが GUI に基づいたものである中で、デバイスを通じてハードウェアの開発に対しても適用できることを実感できるようになっている。また、学習を適宜行う等ある程度対話しているような過程を残すことで既存の LLM サービスのような 1 つのエージェントと対話する形式が保たれており、これは従来の LLM との必要以上の乖離を避け、スムーズな理解を促すことを期待したものである。

### 3.1 インタラクション設計

具体的なフローを図 2 に示す。デバイスは通常モードでは録音と推論を繰り返し、学習モードでは言葉と動作の組み合わせを与えることで新たな内容を学習する。通常モードでは、デバイスは録音を行い、その音声の内容に対して LLM が推論結果を出力し、出力に応じた動作が実行される。録音後には録音結果を文字起こしした内容が画面に表示され、動作中には LLM からの返答である動作内容が画面に表示されている。デバイスは命令に対して基本的な動作単位（例：前進、後退、右/左回り）を組み合わせた一連の動作を行う。

この一連の動作の最中にユーザが物理ボタンを押した場合、デバイスは学習モードに移行する。学習モードでは言葉とそれに対応する出力動作の例をユーザーに指定してもらい、以降デバイスはこの学習内容を考慮した出力を行う。学習モードへの移行後は録音が行われ、ここでユーザは学習させたい言葉を発声する。録音が完了し画面で録音の文字起こし結果を確認し次第、言葉に対応させる動作の入力を行う。プレイマット上でデバイスを手動で動かす

ことでその軌跡が動作として入力される。動かした軌跡は画面に表示されており、ユーザはこれを確認の上、問題がなければボタンを押してデバイスを通常モードに戻す。

### 3.2 実装の詳細

デバイス本体は M5Stack Core2, toio, この 2 つを連結する部分で成り立つ。M5Stack Core2 はマイクロコントローラーモジュール群である。M5Stack 社が開発しもので、ディスプレイや各種センサがあり、このデバイスでは上部に位置している。toio はソニー・インタラクティブエンタテインメントが開発した教育用のロボットプログラミングキットである。Bluetooth を介して M5Stack Core2 と情報の交換が可能である。さらに、専用のパターンが印刷されているプレイマット上であれば、下側面に付いたセンサがマット上の位置を判別できる。Toio はデバイスの下部に配置されている。連結部については、上部で M5Stack Core2 と接続し、下部に toio をはめ込むようなつくりとなっている。連結以外にも、ボタンの追加やマイク、スピーカー、各種センサの拡張などいくつかの機能を M5Stack Core2 に追加するものとなっている。

録音した音声を用いて LLM が出力を行うまでの過程としては、まず音声の文字起こしを行い、その後その文字を LLM に与えて結果を得る。文字起こしには OpenAI が提供する音声認識システムである Whisper の API を利用し、LLM の利用にあたっては OpenAI の API (gpt-4o-2024-05-13) を利用した。どちらも Wifi に接続した M5Stack Core2 上から API にアクセスする。システムプロンプトには





図 3. ユーザー実験の様子。

あなたはロボットであるという旨とどういうロボットであるか、そしてこれまでに与えられた言葉と動作のペアを記述し、推論時は与えられた言葉をそのままユーザープロンプトとして渡している。

## 4 ユーザー実験

本研究の提案するデバイスの評価、また一般ユーザーが LLM を活用するうえで妨げとなっている要素を調査するために、VIVISTOP KANAZAWA<sup>1</sup> のコミュニティに所属する小中学生の参加者を対象に、機械学習を使って遊ぶことを通じて AI への理解を深めるといふ触れ込みでワークショップを開催した。

開催場所は VIVISTOP KANAZAWA のイベントスペースである。ワークショップ冒頭に LLM にまつわる事前知識に関するアンケートを行い、その後実際にデバイスを利用して作品を作り上げる体験や他の参加者の作品に触れる体験を行い、最後に参加者の一部に対してインタビューを行った(図 3)。また、このワークショップは全く同じ内容で 2 回開催された。

### 4.1 事前アンケート

事前知識のアンケートでは一般ユーザーの LLM 活用における課題となりうる、LLM のカスタマイズ性や多機能性に対する認識を調査した。質問を記載した紙を配布し、そこに答える形とした。質問内容としては、これまでに言葉を使う AI を使ったことがあるか、それは何か、を聞いたほか、「言葉を使う AI サービスを活用してこのようなことができると思いますか?」という質問文とともに LLM を利用することができるシチュエーションを複数挙げ、それぞれに対して「思わない」から「思う」までの 5 段階で解答してもらった。

Q1 金沢の歴史について聞く

Q2 なやみごとの相談

Q3 問題を自分らしい文章で回答させる

Q4 今日のことを日記の形で書き起こす

Q5 データをわかりやすくまとめる

Q6 オリジナルのアプリを作る

Q7 ロボットを自在ににあやつる

Q8 機械を設計する

一般ユーザーにとって有名な用途として用意した例が Q1, Q2 である。そして、LLM を自分なりにカスタマイズすることが必要な用途が残りの質問である。多機能性という視点では、Q5, Q6 を比較的複雑なタスクの例として、Q7, Q8 については GUI 以外の用途の例として尋ねた。この選択形式の質問については体験後のアンケートでも同じ質問を行った。

### 4.2 イベント詳細

参加者はまず実験参加の同意を得た上でデバイスの利用方法の説明を受け、その後個人で作品の製作を行い、最後に他の参加者と作品の交換を通じた交流を行った。すべての参加者は同じ会場内に集まり、同じタイムテーブルでワークショップを進行した。実験参加の同意や事前アンケートについてはワークショップの参加受付からワークショップ開始までの保護者が同伴しているタイミングで行い、説明についてはワークショップ開始後スライドを投影しながら一度にまとめて参加者全員に説明を行った。説明内容はこのデバイスは LLM を活用したものである、ということとデバイスの操作方法とそのデモンストレーションのみである。その後、参加者は「遊びを作ってみよう」というテーマのもとで個人で作品を考え、制作した。このようなテーマを与えた理由としては、このデバイスは LLM を使って意外な作品を提案することができることが重要であり、遊びというテーマを与えることは自由に作品を作らせることよりは小中学生に対してそのような体験をより得ることができると判断したためである。さらにオリジナルな遊びを作ることを通じて LLM がそういった自分特有のカスタマイズに対応できることを知ることにつながる。また、作品の交換では作れる作品の幅広さや想像できなかった使い方を知ることによって LLM 自体の可能性の広さを感じることを期待している。

### 4.3 実験後の調査

参加者が作品を交換して交流している中で、数名の参加者に声をかけ、1 人ずつ別室でインタビューを行った。ここではデバイス全体の評価や事前アンケートからの認識の変化に関する聞き取りを行った。詳細の質問内容は会話の流れによって異なるが、満足する作品を作れたか、デバイスは問題なく扱えたか、不満点や改善点はあるか、を中心とした聞き取りを行った。また、アンケートの回答の背景を知る上で、アンケートと同じ質問を混ぜつつ、認識が変化した部分やその理由の聞き取りを行った。

<sup>1</sup> <https://kanazawa.vivita.club>

## 物理的 UI を用いた一般ユーザーの LLM 活用促進手法の提案

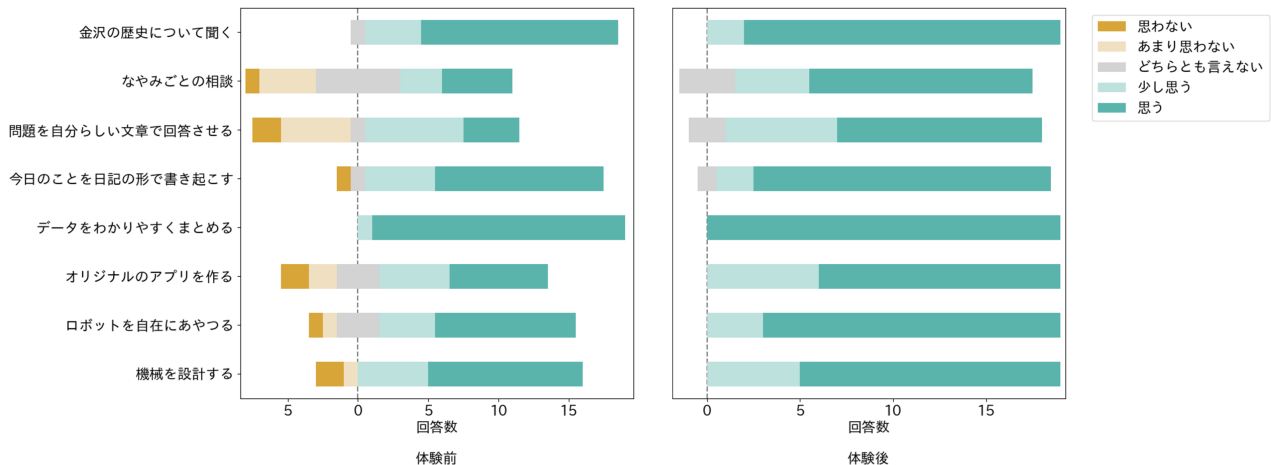


図 4. 体験前後で行なったアンケートの結果。参加者は、「言葉を使う AI サービスを活用してこのようなことができると思えますか?」という問いかけに対し 5 段階のリッカート尺度で答えた。

## 5 結果

### 5.1 参加者の詳細

一日目に 9 人、二日目に 10 人が参加した。参加者の年齢は 9 歳から 14 歳であった。参加者のうち 6 人がこれまでに言葉を使う AI を利用したことがあると回答し、そのうち 4 人は ChatGPT の利用経験があった。

### 5.2 ユーザーの LLM への認識

体験前と体験後に行った 5 段階で答えるアンケートの結果はそれぞれ図 4 のような結果になった。体験前のアンケートの結果から、LLM の事前知識について、挙げたシチュエーションが可能であると既に認識している人が一定数いることがわかる。ただし、シチュエーションの内容によってその内訳は異なっており、Q1、Q5 以外では一定程度の参加者が思わないと答えた。特に Q2、Q3 については実現できると思わない、またはどちらとも言えないとする人の割合が比較的多かった。Q6、Q7、Q8 からは用途の幅広さに対する認識が完全にはなされていないことが伺える。

体験後のアンケート結果と比較した全体の変化としては、質問内容ごとの結果の違いは体験前と大きく変わらないものの、体験前には実現できると思わないと回答された用途があった一方で体験後にはそのような回答が無い。体験前にできると思わない人がいなかった Q1、Q5 においてもより実現できると思う方向に傾いている。

インタビューにおいても、実験の前後で LLM に対する新しい認識が増えたとの回答があった。「課題の相談でしか使っていなかったが、もう少し細かな相談もできそうだった」「画面上の物以外にも影響を及ぼせることに驚いた」といった声があった。

### 5.3 デバイスの使用

デバイスの使用感についてはイベント中の参加者の様子とアンケートの内容から見ていく。最初は自由にデバイスを使ってもらったが、通常モードで動作を確認しつつ学習モードを活用して自分好みの内容を学習させていく、という過程の確認は参加者全員が問題なく行っていた。遊びを作成する段階になってもデバイスの操作については特に問題が見られなかった。作品の一例として、ある参加者はプレイマットに印刷されているマス目を活用し、いくつかのマス目に小さなフィギュアを置きそのマスに到達すると得点を獲得するゲームを提案した。別の参加者は感情ごとに喜びは前、怒りは右と言ったことを学習させ、指示に含まれる感情を読み取り動くデバイスを提案した。

作品を交換する中では、互いに遊びのルールを教え合う中で、相手の遊びを攻略しやすい学習を目指すなど、相手に合わせた LLM のカスタマイズに挑戦できていた。先に挙げたゲームは他の参加者からの人気も高く、ゲームに挑戦した参加者は勝つために効率よくマス目に到達しやすくなることを目指し、網羅的に動くような指示を学習させるなどしていた。インタビューにおいても、「デバイスの使用で不満に思う点はなかった」「特に困らなかった」という意見があり、使いやすさは確保できていたことが伺えた。

### 5.4 ユーザーからの提案

インタビューにおいては、デバイスを利用して感じた欲しい機能、足りない機能を問う質問も行った。「とくに不満には思わなかった」「足りない機能は思いつかない」のように、現状の機能については問題なく受け入れられていた。一方で、「斜めに動きたい」「デバイスのいる場所を覚えてほしい」のよ

うなデバイスの行えることの拡張を望む声は何度か挙げられた。

## 6 議論

### 6.1 考察

参加者の LLM に対する事前の認識としては、アンケートの結果から各用途の実現可能性の認識には幅がある。参加者はおおまかな用途を理解しているが、それが本当に可能かどうかには不安を持っていると考えられる。LLM の使用経験がある参加者が少ない中でも用途に対する認識は一定程度あることから、用途に対する認識があいまいな場合は日常での LLM 活用にあまり繋がっていないことが考えられる。LLM の活用を促していくためには幅広い用途が実現できるという認識をさらに深めていくことが必要条件となってくる。個別の質問で見ると、Q1 の調べものをするのは LLM の用途として高く認知されている。ここからは一般ユーザーにとって単純に知識を LLM から聞くだけのタスクは他に比べ手軽な用途であることが確認できる。一般ユーザーはそのような利用が主となりやすいとも考えられる。一方で、比較的複雑なタスクとして用意した Q5 が可能な用途として高く認知されている。Q5 のデータをまとめるというタスクは、整理の基準や図示の方法などを自身のデータの利用目的に合わせて LLM に与える必要があるという想定であった。しかし、小中学生にとってデータをまとめる行為はデータを決めた時点でほぼ整理の方法が定まっておき、意図と異なる捉え方をされてしまった可能性がある。

デバイスの体験前後での認識の変化の面では、特に LLM の性能に対する認識をあまり持っていなかった参加者に対してデバイスが認識向上の役割を果たしたことが伺える。デバイスから体験できることは言葉を通じてロボットを自在に操ることのみであるが、他の LLM の用途に対しても用途としての期待が持てるようになってきている。

デバイスの操作自体は参加者から不満点が特に得られなかった。これは事前の説明のほかデバイスの画面上に機能を明記していたことが要因であると考えられる。説明では最低限の動作に必要な過程のみを説明していたが、その他の便利機能についてもデバイスを利用する過程で自身で理解し利用している参加者もいた。デバイスの操作方法を自身で深めることができるほどデバイスに対して興味を持たせることができている。物理的な UI と画面による説明がユーザーのスムーズな体験を促していると言える。

デバイスの副次的な効果として遊びの創造や参加者の交流にも有効であることが窺えた。イベント内では参加者に遊びを作るというテーマを与えたが、参加者はそれぞれ独創性の高い遊び提案していた。デバイス単体を利用した遊びから複数人での遊びや

小道具を使用した遊びなどである。ここから、参加者自身がより深く遊びを創造する上でも有用であるとわかる。さらに、参加者が提案した遊びを交換する中で交流を深めることにも繋がっている。

### 6.2 今後の課題

デバイスの機能に対する不満はなかった一方でさらなる機能は求められている。既存のデバイスのままでもプレイマット上の座標を読み取ることや座標を指定した移動などを行うことができ、さらなる動作モジュールの追加も可能である。機能に幅が広がればユーザーはより多彩な表現を学習させることができ、LLM の出力もより複雑なものとなるため、LLM の多機能性の認識という面でも望ましい。

LLM に対する認識についても、LLM ができるかどうか、についての改善は見られたものの、日常での LLM の利用の後押しに繋がるかどうかについてはインタビューの内容から伺えたのみであり明確ではない。認識を改善することはユーザーの LLM 活用を広げていく上での必要な要素ではあるが、十分ではない。LLM のカスタマイズを自分自身で行うことができるかどうかは活用する上では必要な視点であり、そのような点を改善できたか調査する過程も今後必要となる。

参加者は言語モデルが柔軟性の高い出力を行うことに興味を持った一方で、学習を必要としない遊びを提案した参加者もいた。そのような参加者については LLM の出力の多様さ自体には気付いているものの、自分なりの学習を行いその効果を実感するという面では十分とは言えない。学習する過程をより利用されやすくするデバイスや体験の設計をより進めていく必要がある。

また、このデバイスは LLM に慣れ親しんでいないユーザーを対象としているが、イベントと会場の性質上、同世代の中でも相対的に LLM や情報技術に関するある程度の前提知識や興味がある参加者が多かったのも事実である。今後はより幅広い参加者に対して体験を提供、調査していくことが望ましい。

## 7 おわりに

本研究では、一般ユーザーの LLM に対する課題を探ると共に、それを改善するデバイスの提案を行なった。結果として、ユーザーの LLM に対する認識や、デバイスで完結する LLM とのやり取りの効果を伺うことができた。一般ユーザーが自身の LLM 活用の幅を広げられるよう、引き続き LLM の理解促進に向けた取り組みが期待される。

## 謝辞

本研究は JST CREST JPMJCR19F2 の支援を受けたものです。イベントの実施にご協力いただいた

た VIVISTOP KANAZAWA 様, およびクルーの皆様にも深く感謝申し上げます。

## 参考文献

- [1] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.
- [2] M. Community. LLM Survey Report, 2023. Accessed: 2024-08-14.
- [3] S. Diao, P. Wang, Y. Lin, and T. Zhang. Active prompting with chain-of-thought for large language models. *arXiv preprint arXiv:2302.12246*, 2023.
- [4] T. Hitron, Y. Orlev, I. Wald, A. Shamir, H. Erel, and O. Zuckerman. Can Children Understand Machine Learning Concepts? The Effect of Uncovering Black Boxes. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, p. 1–11, New York, NY, USA, 2019. Association for Computing Machinery.
- [5] M. H. Kaspersen, K.-E. K. Bilstrup, and M. G. Petersen. The Machine Learning Machine: A Tangible User Interface for Teaching Machine Learning. In *Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction*, TEI '21, New York, NY, USA, 2021. Association for Computing Machinery.
- [6] T. Kojima, S. S. Gu, M. Reid, Y. Matsuo, and Y. Iwasawa. Large language models are zero-shot reasoners. *Advances in neural information processing systems*, 35:22199–22213, 2022.
- [7] B. Lester, R. Al-Rfou, and N. Constant. The Power of Scale for Parameter-Efficient Prompt Tuning. In M.-F. Moens, X. Huang, L. Specia, and S. W.-t. Yih eds., *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pp. 3045–3059, Online and Punta Cana, Dominican Republic, Nov. 2021. Association for Computational Linguistics.
- [8] X. L. Li and P. Liang. Prefix-Tuning: Optimizing Continuous Prompts for Generation. In C. Zong, F. Xia, W. Li, and R. Navigli eds., *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pp. 4582–4597, Online, Aug. 2021. Association for Computational Linguistics.
- [9] Z. Li, B. Peng, P. He, M. Galley, J. Gao, and X. Yan. Guiding Large Language Models via Directional Stimulus Prompting. In *NeurIPS 2023*, July 2023.
- [10] T. Liu and Y. Sugano. Interactive Machine Learning on Edge Devices With User-in-the-Loop Sample Recommendation. *IEEE Access*, 10:107346–107360, 2022.
- [11] S. Min, X. Lyu, A. Holtzman, M. Artetxe, M. Lewis, H. Hajishirzi, and L. Zettlemoyer. Rethinking the Role of Demonstrations: What Makes In-Context Learning Work? In Y. Goldberg, Z. Kozareva, and Y. Zhang eds., *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pp. 11048–11064, Abu Dhabi, United Arab Emirates, Dec. 2022. Association for Computational Linguistics.
- [12] T. Shin, Y. Razeghi, R. L. Logan IV, E. Wallace, and S. Singh. AutoPrompt: Eliciting Knowledge from Language Models with Automatically Generated Prompts. In B. Webber, T. Cohn, Y. He, and Y. Liu eds., *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 4222–4235, Online, Nov. 2020. Association for Computational Linguistics.
- [13] X. Wang, J. Wei, D. Schuurmans, Q. Le, E. Chi, S. Narang, A. Chowdhery, and D. Zhou. Self-consistency improves chain of thought reasoning in language models. *arXiv preprint arXiv:2203.11171*, 2022.
- [14] J. Wei, X. Wang, D. Schuurmans, M. Bosma, F. Xia, E. Chi, Q. V. Le, D. Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, 35:24824–24837, 2022.
- [15] C. Yang, X. Wang, Y. Lu, H. Liu, Q. V. Le, D. Zhou, and X. Chen. Large language models as optimizers. *arXiv preprint arXiv:2309.03409*, 2023.
- [16] Y. Zhou, A. I. Muresanu, Z. Han, K. Paster, S. Pitis, H. Chan, and J. Ba. Large language models are human-level prompt engineers. *arXiv preprint arXiv:2211.01910*, 2022.
- [17] A. Zimmermann-Niefield, M. Turner, B. Murphy, S. K. Kane, and R. B. Shapiro. Youth Learning Machine Learning through Building Models of Athletic Moves. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children*, IDC '19, p. 121–132, New York, NY, USA, 2019. Association for Computing Machinery.