

DuetML：インタラクティブ機械学習のための ユーザとLLMの協調型フレームワーク

川辺 航^{1,a)} 菅野 裕介^{1,b)}

概要：機械学習（ML）モデルは我々の日常生活の様々な領域に大きな影響を与えている。大規模言語モデル（LLM）は直感的なインターフェースと汎用性を提供する一方で、タスク特化型の軽量モデルは学習や推論の効率性やタスクに特化した性能の面で依然として価値がある。しかし、これらのモデルの開発には技術的な専門知識が必要であり、非専門家のユーザが自身のニーズに合わせてカスタマイズすることは容易ではない。インタラクティブ ML（IML）は直感的な UI を通じて ML 開発の障壁を下げているものの、ユーザにとっては自らの要件を適切なタスクに変換することは難しい。そこで我々は、人間主導の IML と機械主導の LLM アプローチを橋渡しする新しいパラダイムとして、人間と LLM の協調的 ML を提案する。この実現のため、ML プロセス全体でユーザと協働する対話型エージェントとしてマルチモーダル LLM（MLLM）を統合したフレームワーク DuetML を実装した。本システムは、ユーザと MLLM エージェント間の協調的インタラクションを実装することで、MLLM の能力とユーザの主体性の双方を引き出すことを目指す。ユーザ実験を通じて、DuetML がユーザの認知負荷を増加させることなく、ターゲットタスクに適した学習データの定義を可能にし、さらに ML タスクの定式化についてより深い理解を得る機会を提供することを実証した。

1. はじめに

機械学習（ML）モデルは様々な分野で応用され、現代社会の多くの側面に大きな影響を与えている。ML の手法は、SVM [2] や畳み込みニューラルネットワーク（CNN）[6] などの特定タスク向けモデルから、より汎用的な大規模言語モデル（LLM）まで、年々進化を遂げてきた。LLM は多様なタスクに対応可能で直感的な自然言語インターフェースを提供するが、大きな計算資源を必要とする。また、物体検出や画像分類などの特定タスクでは必ずしも優れた性能を発揮せず、非専門家にとって効果的なプロンプトの作成が困難という課題がある。CNN を含む特定タスク向け ML モデルは、計算効率性とタスク特化の性能の面で有用だが、タスクごとの学習が必要で、未知のデータへの対応が困難である。モデルの学習には専門的なプログラミングスキルや数学的知識が必要となり、非専門家ユーザにとっての障壁が高い [13]。

この課題に対し、非専門家によるタスク特化型 ML モデルの開発を容易にするため、インタラクティブ ML（IML）が新たなパラダイムとして研究されてきた [3], [4]。IML

システムは、技術的背景を持たないユーザでも使いやすい UI 操作を通じて ML 開発に参加できるよう設計されてきた。しかし、IML を用いてタスクを適切に解決するモデルを開発することは、必ずしも容易ではないことが報告されている。例えば、ユーザのニーズを適切なタスクに変換することが課題となっている。食用植物を識別するモデルを作成する場合、「トマト」や「キノコ」といった具体的なラベルではなく、「食用」「非食用」「非植物」といった抽象的なカテゴリーを使用の方が望ましいが、初心者ユーザがこのような抽象化に独力でたどり着くことは困難である [11], [13]。

このような障壁をなくすため、本研究では特定タスク向け ML モデル開発の新しいパラダイムとして、人間と LLM の協調的 ML 開発を提案する。IML と LLM は、タスク設定の主導権が人間と機械のどちらにあるかという点で対照的である。人間の知性による完全な制御を目指す IML と、機械知能の汎用化を目指す LLM は、ユーザのエンパワーメントにおいてそれぞれ長所短所がある。我々は、人間と機械が協調的に特定タスク向け ML モデルを作成するプロセスが、非専門家向けの代替解決策となり得ると考える。

本研究では、マルチモーダル LLM（MLLM）の対話機能を活用した、GUI ベースのインタラクティブなフレームワークである DuetML を提案する（図 1）。DuetML は、

¹ 東京大学

^{a)} wkawabe@iis.u-tokyo.ac.jp

^{b)} sugano@iis.u-tokyo.ac.jp

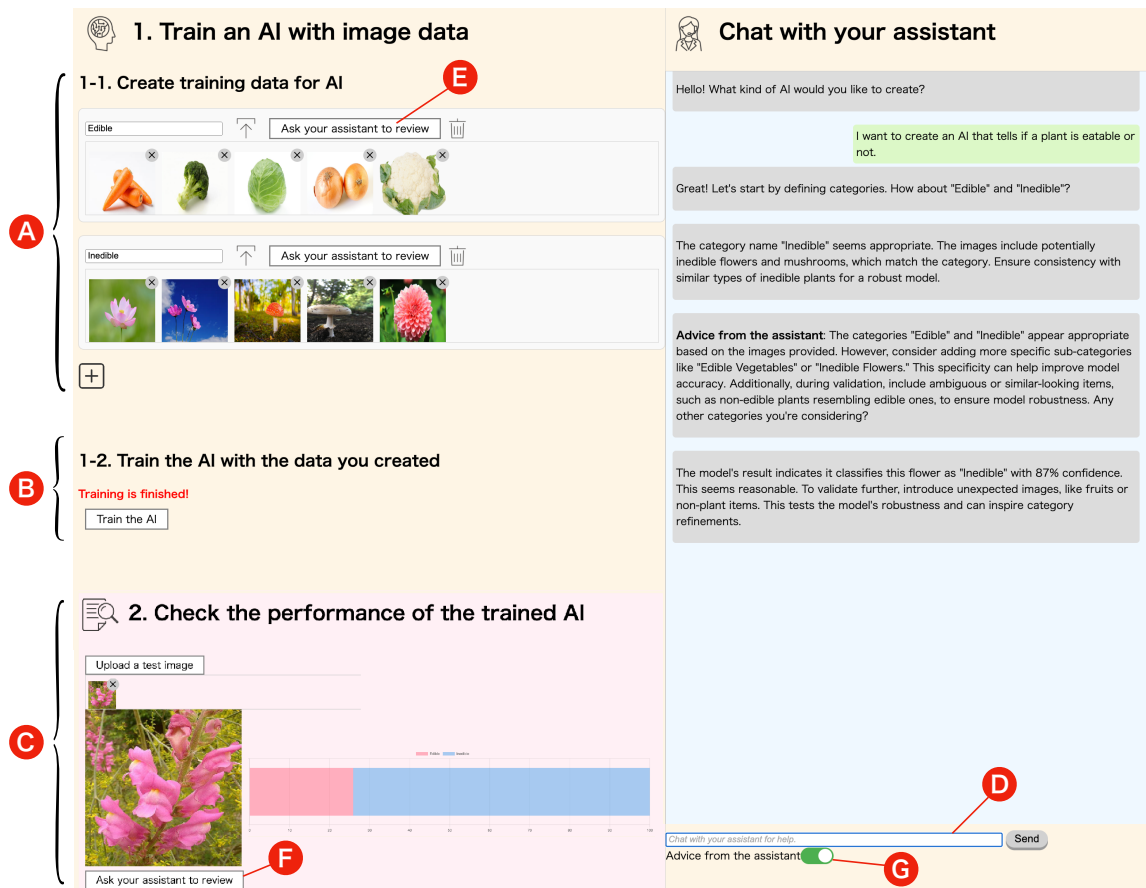


図 1 DuetML の概観.

ユーザの自律性を保ちながら、曖昧な要件を具体的な ML タスクに変換する支援を行う MLLM ベースのエージェントとユーザの協調を促進する。GUI ベースの DuetML は IML スタイルの UI を基盤としており、MLLM ベースのエージェントがユーザの操作履歴、タスクの説明、訓練データの定義を観察しながら、ユーザの作業を支援する。ユーザはエージェントと協力しながら、教師あり学習を通じてモデルの訓練・評価を行う。エージェントによる適切な介入スタイルはユーザごとに異なることを考慮し、ユーザからの要求に応じて反応的に応答するエージェントと、見落とししている可能性のある観点について積極的に提案を行うエージェントの 2 種類を採用した。この動的な協調関係により、ユーザの当初の曖昧なアイデアを、文脈を考慮した対話を通じて適切な訓練データへと変換することを支援する。

我々は、非専門家ユーザを対象とした評価実験を実施し、ML タスク設定におけるエージェントの有効性を検証した。DuetML と協調機能のないベースライン IML システムを比較する方式を採用し、参加者は割り当てられたシステムを使用して、所定の手順に従って画像分類モデルを作成した。結果として、DuetML でエージェントと協調した参加者は、対象タスクにより適合した訓練データを作成できることが示された。参加者は DuetML をベースラインシス

テムより使いやすいとは感じなかったものの、第三者評価からは DuetML がより優れた結果をもたらすと判明した。また、主観評価では、参加者はエージェントとの対話を認知的負荷というよりも有益な機能と捉えていた。これらの知見は、人間と LLM の協調的 ML が ML 開発の民主化に寄与する可能性を示すだけでなく、ML 開発における人間の主体性と機械知能のバランスをどのように取るべきかについての新たな議論へとつながると期待される。

2. 関連研究

2.1 非専門家向けインタラクティブ機械学習

IML 研究の主要な使命は、AI/ML の専門知識を持たないユーザが技術的知識なしで ML モデルを開発できるようにすることである [3], [4]。例えば、非専門家ユーザが直面する技術的課題に対応するための IML ガイドラインの作成 [12] や、二次元平面に埋め込まれたデータをユーザに提示して効果的なインタラクションを促す可視化アプローチ [9], [10] などが研究されてきた。

しかし、ML 開発をより身近にする取り組みが進む一方で、AI/ML リテラシーのないユーザは、自身のニーズを具体的なタスクとして正確に定式化する初期段階で躓くことが多い [11], [13]。従来の IML アプローチは ML の技術的側面をより親しみやすくすることに焦点を当ててき

たが、言語ベースのアプローチはユーザ対話を通じてこの根本的な問題の解決に貢献できる可能性がある。DuetML は、GUI を通じたユーザと MLLM ベースのエージェント間の相互知識交換を促進することで、この課題に取り組んでいる。

2.2 複雑なタスクにおける人間と LLM の対話

HCI 分野では、ユーザによるより効率的で効果的な知的作業実現するため、コンピュータアシスタントや知的エージェントがユーザと協調する文脈が研究されてきた [7], [14]。近年の LLM の進歩により、ユーザの対話を監視し協調する取り組みが、プログラミング支援 [15] や文書作成 [16]、混合現実 [1] など、様々な応用シナリオで実現されている。これらの文脈を考慮した支援の進歩を踏まえ、本研究では MLLM ベースのエージェントがユーザの対話プロセスに介入・協力する新しいパラダイムとして、人間と MLLM の協調的 ML を導入する。エージェントはユーザとの対話履歴や作成された訓練データを分析することで、ユーザの思考プロセスを包括的に把握する。これにより、ユーザの次のアクションを提案し、課題を ML タスクとして定式化する際の議論相手として機能することが可能となる。

3. DuetML

3.1 システム設計

IML システムにおいて非専門家ユーザが直面する課題に対処するため、MLLM を協調的エージェントとして導入した DuetML を開発した。MLLM は視覚的・言語的に状況を理解し、ユーザと協働できる特性を持つものの、MLLM を単独で使用するだけでは軽量の ML モデル開発に適していない。そこで、従来の IML システムの機能を活かしつつ、MLLM をアシスタントとして組み込むことで、ユーザの曖昧なアイデアを具体化する支援を行う設計を採用した。

システムの概要を図 1 に示す。DuetML では、ユーザが ML 開発（教師データの作成 (A) やモデルの訓練 (B)、訓練済みモデルの推論 (C)）を進める過程で、MLLM エージェントがチャットインタフェース (D) を通じて支援を行う。エージェントはユーザの全操作履歴を把握し、適切な介入方法を判断する。システムは 2 種類のエージェントを同時に運用する。ユーザは受動的エージェントに対して教師データや推論結果についての質問ボタン (E および F) で積極的に助言を求めることができ、一方で能動的エージェントは自動的にユーザが見落としがちな観点について助言を生成する。能動的エージェントを煩わしいと思う場合、ユーザはトグルボタンで機能をオフにすることができる (G)。

システムの技術的核心は、視覚情報やユーザの操作状態を含むコンテキストを MLLM に提供する点にある。エージェントは対話履歴や訓練データ画像を含むプロンプト情

報を受け取り、ユーザとの協調的な対話を通じてタスクの定式化を支援する。図 2 に示すように、能動的エージェントと受動的エージェントで異なる処理フローを実装し、それぞれの特性に応じた支援を実現している。

3.2 インタラクションフロー

図 1 に DuetML の GUI を示す。インタフェースは開発領域とチャット領域に分かれている。開発領域では、訓練データの定義 (A)、モデルの訓練 (B)、性能評価 (C) という基本的な ML 開発プロセスを行う。MLLM ベースのエージェントはこの一連の作業を監視し、チャット領域を介して助言を与えるという形で支援を行う。

ユーザが求めるアドバイスの方法の多様性に対応するため、2 つの異なるインタラクションパターンを実装した。1 つ目は受動的エージェントであり、ユーザからの明示的な支援要求に応答する。システム利用開始時、エージェントは「どのような AI を作成したいですか？」と質問し、ユーザはチャット入力 (D) で目標を回答する。この対話を通じてエージェントはユーザのニーズを把握する。その後はチャット入力に加え、訓練データ作成 (E) やモデル評価 (F) の箇所には「アシスタントに質問する」ボタンが配置されており、ユーザは任意のタイミングで受動的エージェントに助言を求めることができる。

自らの理解度を俯瞰的に把握してアドバイスを求めることができるユーザがいる一方で、ユーザ自身はヘルプを求めているもののエージェントが介入した方がより効果的な ML 開発につながるケースがある。このための機能が 2 つ目の能動的エージェントであり、ユーザからの要求なしに定期的にユーザの状態を観測して積極的な助言を行う。例えば、『ニンジン』『花』といった具体的なカテゴリ名を現在定義していますが、『食用植物』『非食用植物』といった広いカテゴリ名の使用を検討してください』といった助言を行う。ユーザは必要に応じてトグルボタン (G) でこの機能をオフにできる。

このように、DuetML は IML の基本機能を保持しながら、MLLM ベースのエージェントによる協調的支援を実現している。これにより、非専門家ユーザが ML モデル開発において直面する課題に対して、より効果的な支援を提供することが可能となっている。

3.3 実装の詳細

DuetML は、ML モデルアーキテクチャ、Web ベースインタフェース、MLLM の統合という 3 つの主要コンポーネントで構成される。分類モデルは、特徴抽出を行う MobileNet [8] と推論を行う SVM [2] を組み合わせた構成を採用した。訓練時には SVM のパラメータのみを更新し、MobileNet のパラメータは固定している。この設計は、最新モデルによる高精度な結果ではなく、訓練プロセスの簡

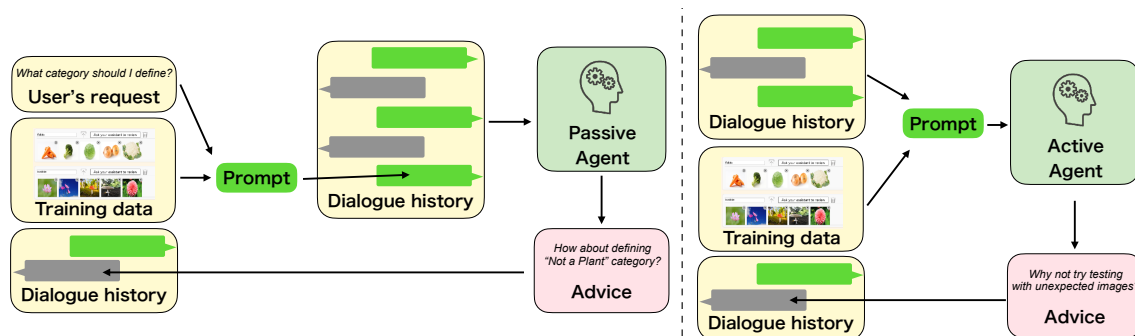


図 2 二種類のエージェントのデータフロー。

潔さと効率性を重視したためである。

インタラクティブな ML 開発を可能にするため、フロントエンドに HTML・jQuery、バックエンドに Python (Flask) を用いた Web ベースインタフェースを実装した。訓練データは各カテゴリごとの画像ファイルとしてサーバーに送信され、グリッドレイアウトで表示される。MLLM の統合は OpenAI API (GPT-4o モデル)*¹を通じて実装し、エージェントごとに異なるプロンプトを採用した。図 2 のようなデータフローで、受動的エージェントはユーザからのリクエストに応じてプロンプトを送信する。また、能動のエージェントは 60 秒間隔で定期的プロンプトを送信する。リクエストの競合を避けるため、2 つの独立した GPT-4o モデルを同時に稼働させている。それぞれのプロンプトは教師データ、ユーザと受動的エージェントとのこれまでの対話履歴、インタラクションログが含まれており、共通する内容としては「ユーザの状況を鑑みてユーザに思考を促すようなアドバイスをしてください」という趣旨が書かれている。

4. ユーザ実験

4.1 実施の概要

DuetML が非専門家ユーザの ML タスク設定を支援できるか、またユーザがシステムをどのように受け止めるかを検証するため、比較ユーザ実験を実施した。ML/AI 経験のない 12 人の非専門家 (平均 34.8 歳の女性 8 名、男性 4 名) を対象とし、研究者が設定した課題に対してモデルの訓練データ作成、訓練、評価を行ってもらった。DuetML との比較対象として、エージェント機能を除いた同一の ML 開発機能を持つベースラインシステムを実装した。公平な比較のため、類似した GUI かつ同一のバックエンドを採用した。

実験は大学の倫理委員会の承認を得て実施し、参加者は大学のメーリングリストやオンライン掲示板で募集した。参加者は無作為に DuetML またはベースラインシステムに割り当てられ、15 分間の練習期間の後、2 つの課題に取

り組んだ。1 つ目は動物画像*²を用いた自由な分類タスク (自由タスク)、2 つ目はさまざまな物体や生き物の画像が入ったデータセット [5] を用いつつ、物体の使用方法に基づく分類という具体的な要件を持つタスクであった (要件タスク)。各課題は最低 10 分、最大 15 分で実施した。

実験後、DuetML を利用した参加者にはエージェントに関する 5 段階のリッカート尺度によるアンケートを実施した。質問項目は以下に示す通りである。

- UQ1 アシスタントからのアドバイスは的確だと感じた、
- UQ2 アシスタントからのアドバイスが役に立った
- UQ3 アシスタントからのアドバイス機能 (= 積極的エージェント) は有益だった

- UQ4 チャットで直接質問する機能は有益だった
- UQ5 ボタンを押して質問する機能は有益だった

また、要件タスクで参加者が作成した訓練データが要件を満たすかを客観的に評価するため、ML 実務経験者 (平均 29.0 歳の女性 1 名、男性 4 名) による第三者評価を行った。評価者は、EQ1 カテゴリ名の適切性、EQ2 画像の妥当性、EQ3 両者の内容的な一致、EQ4 カテゴリの網羅性、EQ5 カテゴリ同士の排他性の観点から、すべての参加者の訓練データを 5 段階のリッカート尺度で評価した。さらに、全参加者に対して、各タスクにどのような方針で臨み、システムとインタラクションする中でそれがどのような変化を受けたかについてインタビューを行った。これについては著者の一人がコーディング含む解析を行った。

4.2 結果

DuetML を用いた参加者によるリッカート尺度評価の結果を図 3 に示す。参加者はエージェントとのインタラクションに関して肯定的な評価を示しており、従来の IML システムに対して追加の要素と言えるエージェントを負担と思わずむしろ有益だと感じていることがわかる。続いて、第三者評価の結果を図 4 に示す。DuetML を使った参加者はベースラインを使った参加者に比べ、ラベル名の定義においてより適切な意思決定ができたことがわかる (マン・

*¹ <https://openai.com/index/hello-gpt-4o/>

*² <https://www.kaggle.com/datasets/iamsouravbanerjee/animal-image-dataset-90-different-animals>

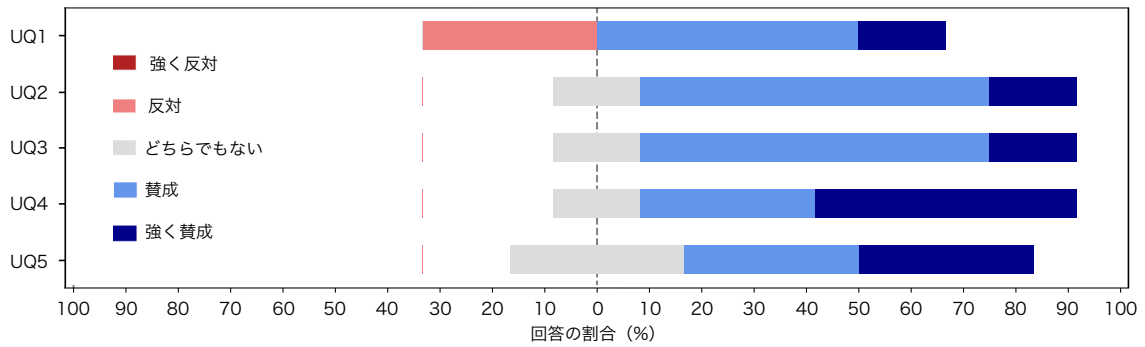


図 3 DuetML のエージェント機能のユーザビリティに関するリッカート尺度評価。

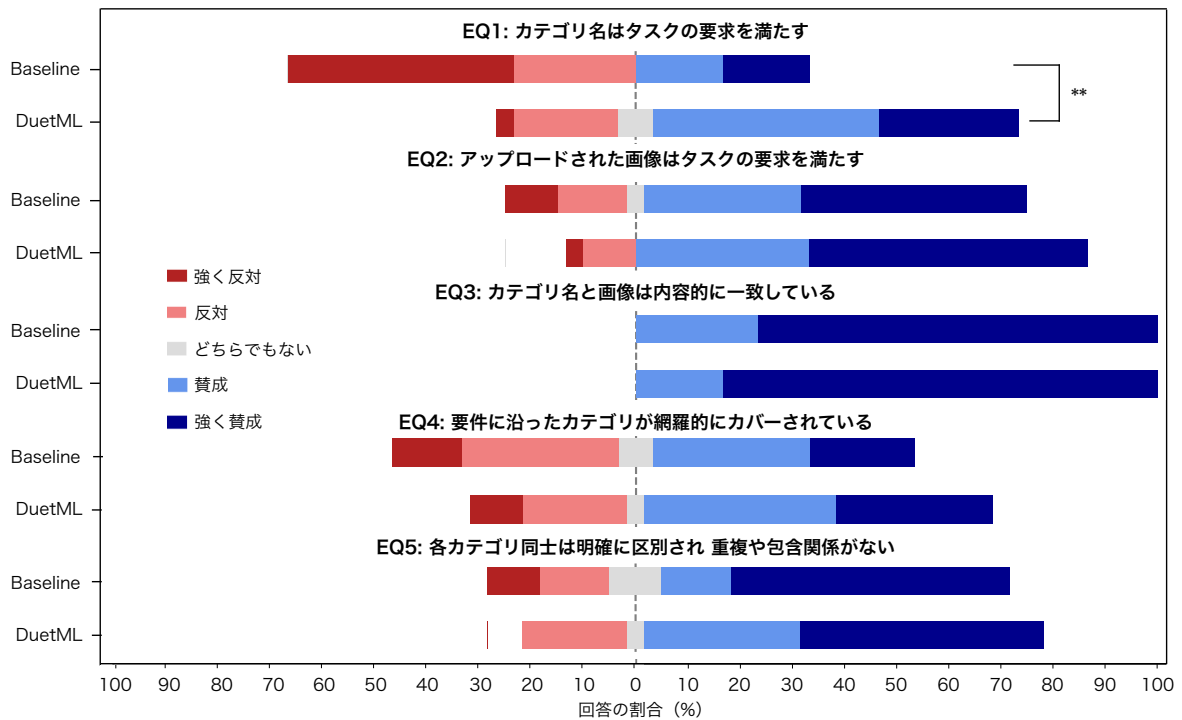


図 4 参加者が作成した教師データに対する第三者評価。

ホイットニーの U 検定にて EQ1 については $p = 0.002$, (ほかは有意差なし)。その他の項目についても DuetML のグループが勝る形となっており, DuetML が適切な教師データ作成を促したことがわかる。

インタビューからは, 参加者が MLLM ベースのエージェントを活用して分類戦略を発展させた 3 つの主要なパターンが明らかになった。

カテゴリの階層の見直し P5 は自由タスクにおいて「哺乳類という大きなカテゴリに多くの画像があったため, エージェントに更なる分類方法を相談した。陸上, 樹上, 水生動物という分類に至り, その妥当性をエージェントと確認しながら進めた」と報告した。

知識の獲得 P10 は「知識が不足している際にエージェントが役立った。蛇が哺乳類か爬虫類かといった質問ができ, 百科事典のように活用できた」と述べた。

抽象的な分類方法の発見 P7 は要件タスクを通じて「エー

ジェントのアドバイスから, 詳細な多数のカテゴリに分けるよりも, 『A と A 以外』のような 2 分類が効果的な場合があることに気付いた」と報告した。

これらのインタラクションパターンから, 参加者がエージェントと積極的に関わりながら, よりニーズに合った教師データを構築していったことがわかる。

5. 考察

5.1 ユーザ実験からの知見

第三者評価の結果から, DuetML を通じた人間と MLLM の協調が, ユーザのタスク定式化と適切な訓練データ作成を効果的に支援することが示された。特にカテゴリ名の定義において DuetML がベースラインシステムを大きく上回る成果を示したことから, ユーザが自らの曖昧なニーズを ML のタスクとして書き起こす過程においてエージェントが効果的に作用したといえる。また, ユーザビリティ評

価からは、MLLM との対話が追加の負担ではなく、むしろ価値ある協調者として認識されていることが明らかになった。これは、MLLM が指示的ではなく、ユーザのアイデアを具体化する対等なパートナーとして機能したためと考えられる。

インタビューからは、DuetML を通じた協調が、ユーザに独自の気づきをもたらし、自らのニーズに対して適切な訓練データとは何かを考えるきっかけを与えたことがわかる。カテゴリの重複や抽象度の適切な設定など、ML タスクの複雑さについて、ユーザは MLLM との対話を通じて深く考察することができた。DuetML の特徴として、MLLM エージェントがトップダウンに正解を与えるに留まらないという点が挙げられる。ユーザの要求に対してチャットボットの的に解を提供するだけでなく、ユーザの現状を踏まえて返答を使い分けることで、ユーザが真に必要とする介入が可能となる。

5.2 今後の課題

DuetML では画像を参照した協調が実現されているが、現状は教師データを参照するために用いられるにとどまっており、今後はよりユーザのコンテキストを深く理解するような使い方が可能である。画像生成やデータ拡張などの技術を組み込むことで、エージェントにとってのより深いユーザ理解が可能となるだろう。理想的には、IML に必要な知識すら必要とせず、単に自然な対話だけでモデルの訓練と評価が可能なシステムを目指すことができる。

また、現在のシステムは ML 開発の基本的な手順に限定されているが、ハイパーパラメータのチューニングやデータ拡張、モデルアーキテクチャの修正など、より複雑な操作もサポートする必要がある。さらに、MLLM の常識的理解を活用してシステムが望ましいモデルの状態を定義し、それに基づいてユーザにデータを要求するような教育的アプローチも検討に値する。これらの課題をクリアすることで、ユーザにとって可能な行動が増え、かつ適切なタスク定義ができる IML システムが実現するだろう。

6. おわりに

本研究では、非専門家ユーザと LLM が協調的に ML 開発を行うためのフレームワークである DuetML を提案した。ユーザと知的エージェントが協調的に ML に取り組むという考え方は画像分類のみならずほかのタスクにも適用可能なものであり、今後の議論の発展が期待される。

参考文献

- [1] Dan Bohus, Sean Andrist, Nick Saw, Ann Paradiso, Ishani Chakraborty, and Mahdi Rad. Sigma: An open-source interactive system for mixed-reality task assistance research—extended abstract. In *2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops*, pp. 889–890. IEEE, 2024.
- [2] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, Vol. 20, No. 3, pp. 273–297, 1995.
- [3] John J Dudley and Per Ola Kristensson. A review of user interface design for interactive machine learning. *ACM Transactions on Interactive Intelligent Systems*, Vol. 8, No. 2, pp. 1–37, 2018.
- [4] Jerry Alan Fails and Dan R Olsen Jr. Interactive machine learning. In *Proceedings of the international conference on Intelligent User Interfaces*, pp. 39–45, 2003.
- [5] Li Fei-Fei, Robert Fergus, and Pietro Perona. One-shot learning of object categories. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 4, pp. 594–611, 2006.
- [6] Jiuxiang Gu, Zhenhua Wang, Jason Kuen, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu, Xingxing Wang, Gang Wang, Jianfei Cai, et al. Recent advances in convolutional neural networks. *Pattern recognition*, Vol. 77, pp. 354–377, 2018.
- [7] Eric Horvitz. Principles of mixed-initiative user interfaces. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pp. 159–166, 1999.
- [8] Andrew G Howard. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.
- [9] Tatsuya Ishibashi, Yuri Nakao, and Yusuke Sugano. Investigating audio data visualization for interactive sound recognition. In *Proceedings of the 25th International Conference on Intelligent User Interfaces*, pp. 67–77, 2020.
- [10] Wataru Kawabe, Yuri Nakao, Akihisa Shitara, and Yusuke Sugano. Technical understanding from interactive machine learning experience: a study through a public event for science museum visitors. *Interacting with Computers*, p. iwae007, 2024.
- [11] Wataru Kawabe and Yusuke Sugano. Image-to-text translation for interactive image recognition: A comparative user study with non-expert users. *Journal of Information Processing*, Vol. 32, pp. 358–368, 2024.
- [12] Swati Mishra and Jeffrey M Rzeszutarski. Designing interactive transfer learning tools for ml non-experts. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–15, 2021.
- [13] Yuri Nakao and Yusuke Sugano. Use of machine learning by non-expert dhh people: Technological understanding and sound perception. In *Proceedings of the Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*, pp. 1–12, 2020.
- [14] Amon Rapp, Lorenzo Curti, and Arianna Boldi. The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots. *International Journal of Human-Computer Studies*, Vol. 151, p. 102630, 2021.
- [15] Steven I Ross, Fernando Martinez, Stephanie Houde, Michael Muller, and Justin D Weisz. The programmer’s assistant: Conversational interaction with a large language model for software development. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*, pp. 491–514, 2023.
- [16] Michele Salvagno, Fabio Silvio Taccone, and Alberto Giovanni Gerli. Can artificial intelligence help for scientific writing? *Critical care*, Vol. 27, No. 1, p. 75, 2023.