卒業論文

LLMマルチエージェント弁証法的推論による質問応答

指導教官　　村上 陽平 教授

立命館大学情報理工学部
先端社会デザインコース

増尾 柚希

2026年1月1日

LLM マルチエージェント弁証法的推論による質問応答

増尾 柚希

内容梗概

大規模言語モデル（LLM）を用いた質問応答システムは, さまざまなアプリケーション, システムなど多様な領域で実用化が進んでいる. 多文化共生社会においては, 質問応答は単なる情報検索にとどまらず, 異なる文化的・宗教的・倫理的背景をもつ人々の相互理解を促進する役割が求められている. 近年,LLM に複数の価値観や立場を付与したマルチエージェント対話が提案され, 模擬的な議論や意思決定を生成する技術が登場している. しかしながら, 既存の LLM ベースのマルチエージェント対話は, 対立する立場をもとに勝敗を決めるディベート的議論が多く, 最終的な結論も表層的な折衷案の提示である. たとえば、「学校清掃を生徒が行うべきか職員が行うべきか」といったトピックにおいて, 現状の LLM は「日常清掃は生徒, 専門清掃は職員」といった妥協案を返す傾向があるが, これは価値の統合ではなく単なる分業の提案にすぎない. 価値観の本質的な対立を踏まえ, 清掃を教育活動として再定義した生徒と職員の共同清掃のような止揚的解決の回答生成は依然として実現できていない. そこで，本研究では, 多様な価値観の対立を扱うトピックを対象に，LLM を用いたマルチエージェント対話によって新たな価値・制度・概念の再構成（止揚）を導く弁証法的な対話プロトコルを提案する. 具体的には，Sawamura, Umeda ら [1] の記号論理に基づく既存の弁証法議論の手続きを参照し，その形式的プロセスを LLM マルチエージェントが自然言語で模倣できるように再構成した段階的対話モデルを設計する. 本手法の実現にあたり，取り組むべき課題は以下の 1 点である. ・記号論理と自然言語の違いに起因する統合プロセスと一貫性保持の困難性弁証法的議論では，各エージェントが立場を一貫して保持しつつ議論を行い，その対立構造を維持した上で，高次の概念として統合（止揚）を生成することが求められる. しかし LLM エージェントは明示的な知識ベースや立場をもたず，対話中に主張が揺らぎやすいため，弁証法議論に必要な立場の一貫性が損なわれやすい. また，統合（止揚）は折衷ではなく価値の再構成による高次な概念の創出を意味するが，その成立条件は自然言語では明確化されておらず，形式的定義だけでは意味的統合として妥当かを判断できない. 本研究では，弁証法議論に必要な「立場の一貫性」を確保するため，まず論証構築フェーズにおいて各

エージェントに主張とその前提を明示的に提示させ，次に反対論証構築フェーズにおいて対立する主張とその前提を提出させるプロセスを設計した. これにより，エージェントの立場と前提が議論開始時点で外部的に固定され，その後の反論・再構成の過程がそれらに基づいて展開されるよう制御した. また，反論段階では，記号論理における攻撃関係の構造を参考に，反論の成立条件を自然言語レベルで再定義し，各発話が初期の立場および提示された前提と矛盾しないよう議論の一貫性を確保した. 統合（止揚）は反論フェーズにて議論を行ったあと、完全に打破されていない、もしくは正当化されていない論証を対象として行うようにした。その際，折衷案とは異なる弁証法的統合を得るために，対立する二つの論証に共通する価値を抽出し，それらをより抽象度の高い上位概念へとまとめ上げる木藤, 栗原ら [2] の一般化（Generalization）の手法を導入した。一般化により，双方の主張を単に接続するのではなく，両者を包摂する新たな概念枠組みを形成しやすくなる。これらの検討結果を踏まえて統合生成用プロンプトを構築し，生成された統合案については定性的評価により，意味的な妥当性や対立包摂の程度を確認した。本研究の貢献は以下の通りである．貢献 1：LLM エージェントを用いて, 弁証法議論を実装したことこれまで記号論理を用いた弁証法議論の形式的定義に関する研究は行われてきたが, 実際に対話トピックを用いて自然言語で対話をさせるものは少なかった. 本研究では, 弁証法議論の形式的定義に準拠しつつ, LLM エージェントを用いて自然言語で対話を行わせる最初の試みになった. 貢献 2：提案プロトコルを用いた弁証法的統合生成の実証実装した LLM マルチエージェント対話システムにより，複数の対立テーマで統合生成を実行し，従来の単なる折衷案では得られない弁証法的統合が生成されることを確認した.

LLM Multi-agent Dialectical Reasoning for Question Answering

Yuzuki MASUO

Abstract

Question answering (QA) systems based on large language models (LLMs) have been widely deployed in various services. In multicultural and value-pluralistic contexts, however, QA is required not only to provide accurate information but also to support consensus building among conflicting viewpoints. Existing LLM multi-agent discussions often converge to debate-style win/lose outcomes or superficial compromises, which makes it difficult to generate an *Aufheben*-like synthesis that reconstructs a higher-level concept acceptable to both sides.

This study proposes a dialectical dialogue protocol in which multiple LLM agents execute argument construction, counterargument (rebut) and premise attack (undercut), iterative defense and refutation, and a final integration phase. Following a phase-controlled procedure inspired by formal computational dialectics, the protocol preserves initial claims and premises to maintain consistency while keeping the conflict structure explicit. In the integration phase, a dedicated agent applies Generalization to extract common grounds and construct a synthesizing argument.

We also discuss practical challenges in translating formally defined dialectical procedures into natural language prompts and in evaluating the quality of synthesized arguments. To address these challenges, we design prompt constraints that fix claims and premises throughout the dialogue and propose an LLM-based evaluation that combines rubric scoring with criteria derived from a formal definition of synthesis.

LLM マルチエージェント弁証法的推論による質問応答

目次

第1章　はじめに

近年，大規模言語モデル（LLM）を用いた質問応答は，検索，学習支援，行政手続き支援など多様なサービスで実用化が進んでいる．一方，多文化共生社会における質問応答には，正確な回答の提示に加えて，異なる価値観の対立を整理し，当事者が受け入れ可能な合意を得るための支援が求められる．学校運営，公共空間の利用，医療・福祉の意思決定などでは，単一の正解を返すこと自体が難しく，複数の立場を踏まえた説明と合意形成の設計が重要となる．

この要請に対し，複数の役割や価値観を付与したLLMエージェント同士の対話により結論を導くマルチエージェント型の枠組みが提案されている．しかし既存手法の多くは，対立する立場をもとに勝敗を決めるディベート的議論，あるいは双方の主張を並列した表層的な折衷案に収束しやすい．その結果，対立を保持したまま共通項を抽出し，新たな価値・制度・概念の再構成として解決を導く止揚的な回答生成には至りにくいという課題がある．

そこで本研究では，弁証法的推論と抽象論証の枠組みに基づき，複数のLLMエージェントが論証構築・反対論証・相互反論・統合を段階的に実行する対話プロトコルを提案する．具体的には，Sawamura らの弁証法議論形式 [1] を参照し，賛成側エージェントが主張と前提列からなる論証を構築し，反対側エージェントが結論否定（rebut）と前提攻撃（undercut）によって攻撃を表明する．反論フェーズでは，攻撃と防御が一定回数まで反復され，いずれかが構築不能と判断した時点で敗北（defeated）または防御可能（defensible）を判定する．最終的に，残存した妥当な主張を対象として，統合エージェントが Generalization[2] により共通項を抽出し，合意可能な止揚論証を構築する．

本研究の実装上の課題は，(1）記号論理で形式定義された弁証法議論と統合メソッドを，自然言語の指示としてLLMに落とし込む難しさと，(2）生成された止揚論証を評価する難しさである．課題（1）に対しては，議論の初期段階で各エージェントに主張と前提を明示的に提出させ，それらを以後の議論で必ず参照させることで，幅広い知識を持つLLM同士でも立場と前提の一貫性を維持するよう制御した．課題（2）に対しては，Anghel らの評価用LLMエージェントによるルーブリック評価 [3] に加え，止揚論証の形式定義をプロンプトとして与えたLLM評価を併用する手法を提案する．

以降では，関連研究と課題設定を整理した上で，提案プロトコルの設計と運

用方法を述べ，評価と考察を通じて有効性と限界を明らかにし，最後に結論と今後の課題をまとめる．

第2章　関連研究

2.1　論証理論に基づく弁証法的議論

対立する主張を形式的に扱う枠組みとして，論証理論および弁証法的議論に関する研究が行われてきた．Sawamura らは，拡張論理プログラミングに基づき，論証の構築，攻撃，防御を段階的に行う Computational Dialectics を提案している [1]．この枠組みでは，結論とその前提列からなる論証を基本単位とし，結論の否定による rebut，および前提の妥当性を否定する undercut によって論証間の攻撃関係を定義する．

また，攻撃と防御の結果として，論証が理論的に破壊された状態（defeated）か，依然として妥当性を保持している状態（defensible）かを判定することで，議論の状態遷移を明確に記述している．このように，勝敗のみで議論を終了させるのではなく，対立が維持された状態を含めて扱える点に特徴がある．

一方で，これらの手法は，有限で明示的な知識ベースを前提とした記号論理的推論を基盤としており，自然言語による柔軟な議論生成や，広範な一般知識を前提とする対話環境への適用は想定されていない．

2.2　Specialization / Generalization に基づく統合手法

弁証法的議論における対立の解決手法として，Kido・Kurihara らは Specialization および Generalization に基づく統合手法を提案している [2]．この枠組みでは，対立する二つの論証を単純に折衷するのではなく，それぞれを限定化した特化（Specialization）や，上位概念へ引き上げる一般化（Generalization）を通じて，両立可能な新たな論証を構築する．

特に Generalization は，対立する主張が共有している前提や価値を抽出し，より抽象度の高い命題として再構成する操作であり，対立を保持したまま合意可能な解を導く 止揚（Aufheben） に対応するものと解釈できる．この点は，双方の主張を部分的に譲歩させる折衷案とは本質的に異なる．

しかし，これらの統合操作も記号論理上の命題操作として定義されており，自然言語による議論において，どのように特化・一般化を実行するかについては明確な指針が与えられていない．

2.3　LLM マルチエージェントによる議論とその限界

近年，大規模言語モデルを用いたマルチエージェント対話による推論や議論生成が注目されている．複数の役割や立場を与えたエージェント同士の対話により，多角的な検討や自己修正が可能となることが報告されている．

しかし，既存の多くの LLM マルチエージェント議論は，ディベート形式による勝敗判定や，複数意見の列挙にとどまり，対立構造そのものを保持したまま統合する機構を持たない．その結果，最終的な回答が表層的な折衷案に収束しやすく，対立から新たな価値や概念を再構成する止揚的な解決には至りにくいという課題がある．

また，生成された議論や結論の評価も困難であり，Anghel らは評価用 LLM エージェントによるルーブリック評価を提案している [3]．この手法は議論の明瞭性や一貫性を評価できる一方で，止揚が成立しているかを直接判定するものではない．

2.4　記号論理と自然言語における議論構造の差異

記号論理に基づく弁証法的議論では，論証は有限で明示的な知識ベースに基づいて構築され，前提，結論，および攻撃関係が形式的に定義されている．このため，どの前提がどの結論を支持しているか，またどの攻撃がどの論証を破壊しているかを一意に特定することが可能である．Sawamura らや Kido・Kurihara らの枠組みは，このような前提の明示性と推論過程の可視性を前提として，弁証法的議論や統合操作を定式化している．

一方，大規模言語モデルを用いた自然言語による議論では，推論は暗黙的な背景知識や文脈理解に強く依存しており，論証の前提や境界が明示されないまま議論が進行することが多い．LLM は豊富な一般知識を利用できる反面，議論の過程で新たな前提や視点を自律的に導入しやすく，どの知識に基づいて主張がなされたのかを外部から厳密に把握することが困難である．

この差異は，議論の進行様式にも影響を与える．記号論理に基づく議論では，攻撃や防御はあらかじめ定義された論証や前提を対象として行われるため，議論は閉じた構造の中で推移する．これに対し，自然言語による議論では，話題の拡散や論点のすり替えが生じやすく，攻撃と防御の対応関係が不明確になりやすい．その結果，対立構造を保持したまま議論を収束させることが難しく，最

終的な結論が表層的な折衷や意見の併記にとどまる傾向がある．

このように，記号論理に基づく弁証法的議論の形式的厳密さと，自然言語・LLM による議論の柔軟性との間には本質的な隔たりが存在する．そのため，既存の論理ベースの弁証法議論や統合手法を，そのまま自然言語・LLM 環境に適用することは困難であり，両者の特性を踏まえた新たな設計が必要となる．

2.5 本研究の位置づけ

以上の関連研究を踏まえると，記号論理に基づく弁証法的議論および統合手法は，対立構造の厳密な定義という強みを持つ一方で，自然言語・LLM 環境への適用が困難である．一方，LLM マルチエージェント議論は柔軟な対話生成が可能であるものの，止揚に至るための理論的枠組みを欠いている．

本研究は，これら二つの研究潮流を接続し，弁証法的議論と抽象論証の形式的枠組みを参照しつつ，それらを自然言語による LLM エージェント対話として再設計する点に特徴がある．特に，rebut ／ undercut に基づく攻撃・防御，defeated ／ defensible の状態判定，および Generalization による統合を一貫した対話プロトコルとして統合し，止揚的回答生成を可能とする点で既存研究と異なる位置づけにある．

第3章　提案手法

3.1　設計方針と全体構成

前章で述べたように，記号論理に基づく弁証法的議論と，自然言語・LLM による議論との間には，前提の明示性や議論構造の閉性において本質的な差異が存在する．本研究では，この差異を踏まえ，論理ベースの弁証法議論が持つ構造的厳密さを維持しつつ，LLM による柔軟な自然言語議論を可能とする対話プロトコルを設計する．

本研究におけるエージェントとは，大規模言語モデルに対して特定の役割と推論制約を与え，論証構築，攻撃，防御，および統合といった操作を担当させた対話主体を指す．提案プロトコルでは，賛成側エージェント（AG1），反対側エージェント（AG2），統合エージェント（AG3）の三者を用いる．

プロトコル全体はフェーズ管理に基づいて実行され，各フェーズで参照可能な情報を明示的に制限することで，議論の一貫性と停止性を保証する．図 1 に提案プロトコルの全体構成を示す.

3.2　弁証法的対話プロトコル

本研究で提案する弁証法的対話プロトコルは，複数の LLM エージェントが役割に応じて発話を行い，論証の構築，攻撃，防御，および統合を段階的に実行する枠組みである．プロトコルはフェーズ管理に基づいて制御され，各フェーズで参照可能な情報を明示的に制限することで，議論の一貫性と停止性を保証する．図 1 にプロトコルの全体構造を示す.

以下では，プロトコルを構成する各フェーズを順に説明する．

3.2.1　論証構築フェーズ（AG1）

賛成側エージェント AG1 は，対象となる問いに対して，自身の立場を表す主張と，それを支持する前提列からなる論証を構築する．この初期論証は，議論全体の基準点として位置づけられ，以後のフェーズでは，この主張および前提のみが議論対象として参照される．

この設計は，LLM が推論過程で新たな前提や論点を自律的に導入することを防ぎ，記号論理における有限知識ベースを，自然言語環境において近似的に再現することを目的としている．

3.2.2　反対論証・攻撃フェーズ（AG2）

反対側エージェント AG2 は，AG1 の主張に対して結論の否定による反対論証（rebut）を構築し，立場の対立を明示する．続いて AG2 は，AG1 の前提に対して undercut を用いた攻撃を構築する．

各 undercut は，対象とする前提を明示的に指定した攻撃ノードとして管理され，後続の反論フェーズにおいて独立に扱われる．これにより，議論の焦点が特定の前提に固定され，話題の逸脱を防ぐ．

3.2.3　相互反論フェーズ

相互反論フェーズでは，AG2 が構築した各攻撃ノードに対し，AG1 が防御可能かを判断する．防御が可能と判断された場合，AG1 は対応する反論を構築する．

防御を受けた AG2 は，その反論に対して再反論が可能かを判断し，可能な場合のみ反論を行う．これらの攻撃と防御は，あらかじめ定めた回数に達するまで反復される．この反復制御により，議論は有限ステップで必ず停止する．

3.2.4　敗北判定フェーズ

相互反論フェーズ終了後，プロトコルは議論の状態を判定する．AG1 の論証が理論的に破壊された場合，AG1 は敗北（defeated）と判定される．一方，AG2 の全ての攻撃が防御され，かつ反論不能となった場合，AG2 が敗北と判定される．

いずれの敗北条件にも該当しない場合，論証は防御可能状態（defensible）と判定される．この状態は，対立の妥当性が保持されているが，いずれの立場も決定的に否定されていない状況を表す．

3.2.5　統合（止揚）フェーズ（AG3）

統合フェーズでは，統合エージェント AG3 が，防御可能と判定された論証および反対論証を入力として受け取り，止揚論証を構築する．本研究では，Kido・Kurihara らの Generalization を参照し，対立する主張が共有している前提や価値を抽出し，より抽象度の高い合意可能な命題として再構成する．

この統合操作は，双方の主張を部分的に譲歩させる折衷ではなく，対立を保持したまま新たな概念や制度を導出する止揚として位置づけられる．

3.3 小結

本章では，弁証法的推論に基づく LLM マルチエージェント対話プロトコルを提案し，その設計方針および各フェーズの処理内容を示した．次章では，生成された止揚論証を評価するための手法について述べる．

第4章　評価手法

4.1　評価の課題と方針

本研究が対象とする止揚的回答生成は，単一の正解や最適解を持たない問題を扱う．そのため，生成結果の妥当性を自動的かつ一意に判定することは困難であり，従来の正解一致率やスコアベースの評価指標をそのまま適用することはできない．

特に，本研究では，対立する立場を保持したまま共通項を抽出し，新たな価値・概念・制度を再構成することを目的としている．このような止揚論証の成立は，単なる情報網羅性や論理的一貫性のみでは評価できず，対立の包摂性や再構成の創発性といった観点を含めて判断する必要がある．

以上を踏まえ，本研究では，評価用 LLM エージェントによるルーブリック評価と，止揚論証の形式定義に基づく評価を組み合わせた定性的評価手法を採用する．

4.2　評価対象と実験設定

評価対象は，第 3 章で提案した弁証法的対話プロトコルにより生成された最終回答，すなわち統合フェーズにおいて構築された止揚論証である．比較対象として，既存の LLM マルチエージェントによる議論生成手法，および単一 LLM による多段推論結果を用いる．

評価は，同一の問いに対して各手法が生成した回答を入力とし，評価用 LLM エージェントがそれらを独立に判定する形で実施する．評価者には議論過程や内部プロンプトを与えず，最終的に生成された回答文のみを提示することで，回答としての受容可能性に焦点を当てた評価を行う．

4.3　ルーブリックに基づく評価

4.3.1　評価指標

ルーブリック評価では，Anghel ら [3] の手法を参照し，以下の観点から止揚論証を評価する．

明瞭性（Clarity）回答が構造的に整理され，読み手にとって理解しやすいか．

一貫性（Coherence）主張とその根拠が矛盾なく接続されているか．

対立包摂性（Inclusiveness）対立する立場の主張が公平に取り込まれているか．

止揚の創発性（Novelty of Synthesis）単なる折衷や要約ではなく，新たな観点や概念が提示されているか.

各項目は段階評価とし，評価用 LLM エージェントが理由付きでスコアを付与する.

4.3.2　評価用 LLM エージェント

評価には，議論生成に用いたモデルとは独立した LLM を用い，評価専用のプロンプトを与える．これにより，生成モデル自身による自己評価を避け，評価の客観性を高める.

評価用プロンプトでは，各評価項目の定義を明示し，生成された回答がそれぞれの観点をどの程度満たしているかを文章で説明させた上でスコアを算出する.

4.4　止揚論証の形式定義に基づく評価

評価には，議論生成に用いたモデルとは独立した LLM を用い，評価専用のプロンプトを与える．これにより，生成モデル自身による自己評価を避け，評価の客観性を高める.

評価用プロンプトでは，各評価項目の定義を明示し，生成された回答がそれぞれの観点をどの程度満たしているかを文章で説明させた上でスコアを算出する.

4.4.1　止揚の成立条件

ルーブリック評価に加え，本研究では，止揚論証の成立をより厳密に検証するため，形式定義に基づく評価を行う．本研究では，対立する二つの主張 A，B に対して，以下の条件を満たす命題 C を止揚論証とみなす.

C は A および B のいずれか一方のみを単純に肯定・否定するものではない.

C は A および B が共有する前提や価値を明示的に含む.

C は A，B の同時成立を前提としない新たな視点を導入している.

これらの条件は，記号論理における Generalization に基づく統合条件を，自然言語議論に適用可能な形に再解釈したものである.

4.4.2　形式定義を用いた LLM 評価

形式定義に基づく評価では，上記の成立条件をプロンプトとして評価用 LLM に与え，生成された回答が各条件を満たしているかを判定させる．この際，単なる Yes/No 判定ではなく，各条件に対する満足度とその根拠を文章で説明させる.

この評価により，回答が折衷案や意見の列挙にとどまっていないか，止揚と

して成立しているかを定性的に検証する．

4.5　評価手法の位置づけ

本研究の評価手法は，単一の数値指標による自動評価ではなく，複数の観点から止揚論証の質を検証する定性的評価を採用している点に特徴がある．ルーブリック評価は回答としての読みやすさや説得力を捉え，形式定義に基づく評価は止揚としての理論的妥当性を検証する．両者を併用することで，止揚的回答生成の有効性を多面的に評価することが可能となる．

第5章　評価結果と考察

5.1　実験結果の概要

本節では，第4章で定義した評価手法に基づき，提案手法および比較手法によって生成された回答の評価結果を概観する．各評価観点において，提案手法は，既存の LLM マルチエージェント議論および単一 LLM による推論と比較して，対立構造の明示性および統合結果の一貫性において特徴的な傾向を示した．

以下では，評価観点ごとに結果を詳細に分析し，その要因について考察する．

5.2　ルーブリック評価結果

5.2.1　明瞭性・一貫性に関する結果

明瞭性および一貫性の観点では，提案手法による回答は，議論の流れが構造的に整理されている点で高い評価を示す傾向が見られた．これは，論証構築，反対論証，相互反論，統合というフェーズを明示的に分離したことにより，主張，対立，結論が混在しにくくなったためと考えられる．

5.2.2　対立包摂性に関する結果

対立包摂性の観点では，提案手法は，一方の立場のみを強調する回答や，反対意見を形式的に列挙するだけの回答を避ける傾向を示した．これは，敗北判定フェーズにおいて，いずれかの立場を早期に排除するのではなく，防御可能状態（defensible）を維持した論証を統合対象とした設計による影響であると考えられる．

5.2.3　止揚の創発性に関する結果

止揚の創発性に関しては，提案手法による回答が，単なる折衷案や中立的要約にとどまらず，新たな観点や概念的枠組みを提示する例が確認された．これは，統合フェーズを議論フェーズから独立させ，統合専用のエージェントに処理を委ねたことが寄与していると考えられる．

5.3　形式定義に基づく評価結果

5.3.1　止揚成立条件の満足度

止揚論証の形式定義に基づく評価では，提案手法による回答が，対立する主張のいずれかを単純に肯定・否定する形になっていないこと，および両者が共有する前提や価値を含んでいることが確認された．一方で，統合結果の抽象度

や新規性にはトピック依存のばらつきが見られた．

5.3.2　折衷案との差異

比較手法では，対立する主張の要点を併記したり，中間的立場を提示する折衷的回答が多く見られたのに対し，提案手法では，対立構造を再構成する形での回答生成が行われる傾向が確認された．この差異は，統合を最終フェーズとして明示的に位置づけた点に起因すると考えられる．

5.4　総合考察

本研究の結果から，提案手法の有効性は，主に以下の三点に集約される．

第一に，記号論理に基づく弁証法的議論が持つ前提の明示性や攻撃構造を，自然言語・LLM 環境において近似的に再現した点である．前提固定や攻撃対象の限定により，議論の拡散が抑制され，対立構造が保持された．

第二に，LLM マルチエージェントにおいて役割とフェーズを明確に分離した点である．賛成，反対，統合という役割分担と，段階的なフェーズ管理により，各エージェントの推論目的が明確化され，議論が手続き的に制御された．

第三に，止揚を独立した生成タスクとして扱い，評価可能な対象として位置づけた点である．これにより，合意形成型 QA における止揚的回答生成の実現可能性が示された．

5.5　限界と課題

本研究では，評価を主に定性的手法に依存しており，評価用 LLM に内在するバイアスの影響を完全には排除できていない．また，フェーズ数や反論回数の設定は経験的に定めたものであり，議論の複雑性に応じた最適化が今後の課題である．

さらに，扱うトピックの範囲や実運用環境における有効性については，今後の検証が必要である．

第6章　おわりに

本研究では，形式定義された弁証法議論と統合メソッドに基づき，LLM マルチエージェントが合意可能な止揚論証を生成するための対話プロトコルを提案した．提案プロトコルは，論証構築から攻撃・防御・相互反論を段階的に実行し，残存する妥当な主張を対象として Generalization による統合を行う．評価では，対話の明瞭性と一貫性，および止揚の創発性に関して，既存の手法と比較して有効性が示唆された．

今後の課題として，評価の再現性向上に向けた人手評価との併用，反論回数や停止条件の最適化，および実運用に近い条件での検証が挙げられる．

謝辞

本研究の遂行にあたり，ご指導・ご助言を賜った関係者の皆様に深く感謝する．

参考文献

[1] Sawamura, H. and Umeda, Y.: Computational Dialectics for Argument-based Agent Systems, *Proc. ICMAS 2000*, pp. 271–278 (2000).

[2] Kido, H. and Kurihara, M.: Computational Dialectics Based on Specialization and Generalization, *Proc. JSAI 2008*, pp. 228–241 (2008).

[3] Anghel, C. et al.: Multi-Model Dialectical Evaluation of LLM Reasoning Chains, *Informatics*, Vol. 12, No. 3, p. Art. 76 (2025).