様式７号（第 12 条，第 31 条関係）
（様式７号）(Format No.7)　日本語版

学　位　論　文　要　旨
(Summary of the Doctoral Dissertation)

学位論文題目 (Dissertation Title)	Prompt-Guided Navigation in the Reasoning Space of Large Language Models (大規模言語モデルの推論空間におけるプロンプト誘導型ナビゲーション)
氏　名(Name)	ZHU MENGYAO

本博士学位論文は，数種類のプロンプト方略が，大規模言語モデル（LLM）による高品質な数学的定義の生成能力にどのように寄与するかを体系的に検討するものである。数学的定義の生成は，大学レベルの数学学習において人工知能にとって基礎的でありながら認知的に困難な課題である。3 つの連続した研究を通じて，本論文は，推論の構造（線型推論および反復的推論）と文脈規定の精緻度（細分化された教授認識型ペルソナ割り当て）の両方が，LLM がその内部的推論空間をどのように探索して高品質な出力を生み出すかを根本的に規定することを実証する。新たに考案し統計的に検証した 3 つのプロンプト方略，すなわち Cumulative Thinking Prompt，AUDIT Prompt，および Professor-Rank-Aware Prompt を導入することにより，本論文は，知識集約的タスクにおけるモデルの性能向上のための，エビデンスに基づいた学習不要の手法を提供し，教育および技術的分野への直接的な応用可能性を持つとともに，STEM 分野の数学学習およびチュータリングに対して正の社会的影響をもたらす。本論文は 6 つの構成章からなり，以下にその概要を述べる。

第 1 章では，研究の動機，目的，および論文の構成を紹介し，数学的定義の生成を，LLM の膨大な参数的知識空間内における推論ナビゲーション問題として位置付ける。既存のプロンプト方略は，定義生成を，構造化された知識の蓄積，反復的洗練，または最適化されたペルソナ指定のメカニズムを欠いた一回限りのプロセスとして扱う傾向があり，その点に重大なギャップが存在することを指摘する。続いて，3 つの逐次的な研究目的を概説する。(1) 生成前の知識蓄積を伴う線型推論手法（Cumulative Thinking Prompt）の開発，(2) 再帰的品質監査を組み込んだ反復的推論フレームワーク（AUDIT Prompt）の設計，(3) 細分化されたペルソナ指定と粗いペルソナ指定の影響（Professor-Rank-Aware Prompt）の検討である。最後に，対処した主要な課題として，18 の数学分野にわたる大規模でバランスの取れたデータセットの構築（研究 1 では 439 項目，研究 2 および 3 では 921 項目に拡張）と，検証済みの多指標自動評価フレームワークについて述べる。

第 2 章では，Cumulative Thinking Prompt の研究を提示する。この研究は，最終的な数学的定義を生成する前に，LLM に対して本質的理解から抽象的理解へと進む 5 段階の逐次的な知識想起ステップを実行するよう指示する線型推論手法を導入する。439 の数学用語に対して評価し，3 つのベースライン（Direct Prompt，Excellent-Professor-Aware Prompt，Step-by-Step Prompt）と比較した結果，Cumulative Thinking Prompt は最高の性能を示し，最低性能のベースラインに対して平均総合相対スコア 104.11%を達成した。本章は，生成前の構造化された知識蓄積が効果的な認知的足場として機能し，モデルを最適な推論部分空間へと導き，意図的な準備的推論が定義の質に測定可能な改善をもたらすことを結論付ける。

第 3 章では，生成プロセスに再帰的品質監査と洗練サイクルを組み込んだ反復的推論フレームワークである AUDIT Prompt を提案する。線型推論手法とは異なり，このアプローチは，初期定義を生成し，構造化された

(和文 2,000 字程度 / 英文 800 語程度)
(about 800 words)

品質監査を実施し，洗練された定義を生成し，2 回目の監査を実施し，最終的な洗練定義を出力する。921 の数学用語からなる拡張データセットで評価した結果，AUDIT V3 はベースライン（Excellent-Professor-Aware Prompt）に対して総合相対性能 114.42%を達成し，Cumulative Thinking Prompt を 11.61 パーセントポイント上回るとともに，卓越した安定性（標準偏差 0.058）を示した。本章は，明示的なフィードバック機構を備えた反復的推論が推論経路の動的調整を可能にし，形式的厳密性と概念的完全性が要求されるタスクにおいて，線型推論アプローチよりも大幅に高くかつ一貫した品質をもたらすことを確立する。

第 4 章では，Professor-Rank-Aware Prompt フレームワークを通じてペルソナの特異性の役割を検討し，Distinguished Professor から Lecturer に至る 8 段階の細分化されたアカデミックランク指定を調査する。結果は，アカデミックな年次と一致する明確な性能階層を示した。すなわち，Distinguished Professor は相対性能 106.3%，Emeritus Professor は 103.8%，Full Professor は 103.1%，Associate Professor は 99.4%，Teaching Professor は 99.0%，Assistant Professor は 97.4%，Lecturer は 97.2%であり，いずれも粗い「Excellent Professor」ベースラインに対して基準化された。重要なことに，4 つの上位ランクのバリエーションはベースラインを上回り，4 つの下位ランクのバリエーションは下回り，一般化された未検証のペルソナ設定が最適ではないことを確認した。本章は，精緻なペルソナ指定が効果的なナビゲーション信号として機能し，モデルをより価値ある数学的情報を含む推論部分空間へと導くことを結論付ける。

第 5 章では，3 つの研究全体にわたる知見を統合し，達成された点と改善すべき点を論じる。Cumulative Thinking Prompt は，線型的で構造化された知識蓄積が定義の質を高めることを確立し，AUDIT Prompt は，再帰的洗練を伴う反復的推論がより優れた安定した性能をもたらすことを示し，Professor-Rank-Aware Prompt は，細分化されたペルソナ指定がモデルの推論空間のナビゲーションを根本的に形作ることを明らかにした。本章は，反復的洗練サイクルの計算コスト，適応的終了戦略の必要性，多様な学習者集団に対する教育学的明瞭さを完全には捉えきれない可能性のある形式的数学的品質指標への焦点，およびペルソナ研究で使用された西洋中心のアカデミックランクなど，主要な限界を認識する。これらの限界は，今後の研究において方法論的洗練を必要とする具体的な領域を示している。

第 6 章では，今後の方向性を探求し，3 つの研究方法論の相乗的統合と動的選択を構想する。期待される進展としては，線型推論と反復的推論を文脈認識型ペルソナ選択と組み合わせ，タスクの複雑性とリアルタイム品質評価に基づいて方略を適応させるハイブリッドフレームワークが挙げられる。別の方向性として，特殊化された推論行動を持つ異なるペルソナが協働して推論空間をナビゲートするマルチエージェント協調が考えられる。教授ランクフレームワークは多様な学習者集団に拡張されうる。最後に，本研究を通じて開発された体系的方法論は，科学的執筆，コード生成，法的文書作成，医療レポート合成など，精緻で構造化された出力を必要とする他の分野において LLM の振る舞いを最適化するためのテンプレートを提供する。

要約すると，本論文は，知識集約的タスクに対するプロンプトエンジニアリングに対し，3 つの相互に関連する社会的貢献を行う。すなわち，生成前の構造化された知識蓄積を伴う線型推論（Cumulative Thinking Prompt），再帰的品質監査を伴う反復的推論（AUDIT Prompt），および細分化されたペルソナ指定（Professor-Rank-Aware Prompt）のそれぞれが，数学的定義生成に対して明確かつ実証的に検証された利点を提供することを示す。総合的に，本研究の知見は，意図的なプロンプト設計，具体的には推論の軌跡と文

脈の精緻度の制御が，大規模言語モデルから高品質な出力を引き出すための信頼性の高い学習不要の経路を提供し，教育技術および自動コンテンツ生成に重要な示唆を与え，それによって正の社会的影響をもたらすことを確立する。

様式 7 号（第 12 条，第 31 条関係）
（様式 7 号）(Format No.7) 英語版

学 位 論 文 要 旨
(Summary of the Doctoral Dissertation)

学位論文題目 (Dissertation Title)	Prompt-Guided Navigation in the Reasoning Space of Large Language Models (大規模言語モデルの推論空間におけるプロンプト誘導型ナビゲーション)
氏　名(Name)	ZHU MENGYAO

This doctoral dissertation systematically investigates how distinct prompt strategies enhance the large language model's abilities in generating high-quality mathematical definitions, a foundational yet cognitively demanding task in college-level math learning for artificial intelligence. Across three sequential studies, the research establishes that both the structure of reasoning (linear reasoning and iterative reasoning) and the precision of contextual specifications (fine-grained professor-aware-persona assignments) fundamentally shape how an LLM navigates its internal reasoning space to produce high quality outputs. By introducing and statistically validating three novel prompt strategies, the Cumulative Thinking Prompt, the AUDIT Prompt, and the Professor-Rank-Aware Prompt, the dissertation provides evidence-based, training-free approaches for enhancing model performance on knowledge-intensive tasks, with direct applications in educational and technical domains and contribute positive social impacts on STEM math learning and tutoring. The dissertation includes six component chapters which will be introduced in the following paragraphs.

Chapter 1 introduces the motivation, objectives, and dissertation outline, framing mathematical definition generation as a reasoning navigation problem within the vast parametric knowledge space of LLMs. It identifies a critical gap in existing prompt strategies, which typically treat definition generation as a one-pass process lacking mechanisms for structured knowledge accumulation, iterative refinement, or optimized persona specification. The chapter then outlines three sequential research objectives: (1) developing a linear reasoning method with pre-generation knowledge accumulation (Cumulative Thinking Prompt), (2) designing an iterative reasoning framework with recursive quality audits (AUDIT Prompt), and (3) investigating the impact of fine-grained versus coarse-grained persona specifications (Professor-Rank-Aware Prompt). Finally, it describes key challenges addressed, including the construction of a large, balanced dataset spanning 18 mathematical subjects (439 terms in Study 1, expanded to 921 terms in Studies 2 and 3) and a validated, multi-metric automatic evaluation framework.

Chapter 2 presents the Cumulative Thinking Prompt study, which introduces a linear reasoning method that instructs the LLM to perform five sequential knowledge-recalling steps, progressing

from essential to abstract understanding, before generating a final mathematical definition. Evaluated across 439 mathematical terms and compared against three baselines (Direct Prompt, Excellent-Professor-Aware Prompt, and Step-by-Step Prompt), the Cumulative Thinking Prompt achieved superior performance, with an average holistic relative score of 104.11% against the lowest-performing baseline. The chapter concludes that structured knowledge accumulation prior to generation serves as an effective cognitive scaffold, guiding the model toward optimal reasoning subspaces and demonstrating that deliberate preparatory reasoning yields measurable improvements in definition quality.

Chapter 3 proposes the AUDIT Prompt, an iterative reasoning framework that incorporates recursive quality audits and refinement cycles into the generation process. Unlike linear reasoning methods, this approach generates an initial definition, conducts a structured quality audit, produces a refined definition, performs a second audit, and delivers a final refined definition. Evaluated on an expanded dataset of 921 mathematical terms, AUDIT V3 achieved a holistic relative performance of 114.42% against the baseline (Excellent-Professor-Aware Prompt), outperforming the Cumulative Thinking Prompt by 11.61 percentage points while demonstrating exceptional stability (standard deviation of 0.058). The chapter establishes that iterative reasoning with explicit feedback mechanisms enables dynamic adjustment of reasoning paths, producing substantially higher and more consistent quality than linear reasoning approaches for tasks requiring formal precision and conceptual completeness.

Chapter 4 investigates the role of persona specificity through the Professor-Rank-Aware Prompt framework, examining eight fine-grained academic rank specifications from Distinguished Professor to Lecturer. The results reveal a clear performance hierarchy aligned with academic seniority: Distinguished Professor achieved 106.3% relative performance, Emeritus Professor 103.8%, Full Professor 103.1%, Associate Professor 99.4%, Teaching Professor 99.0%, Assistant Professor 97.4%, and Lecturer 97.2%, benchmarked against a coarse-grained "Excellent Professor" baseline. Critically, four higher-ranked variations outperformed the baseline, while four lower-ranked variations underperformed, confirming that general, untested persona settings are suboptimal. The chapter concludes that precise persona specifications serve as effective navigation signals, guiding models toward reasoning subspaces containing more valuable mathematical information.

Chapter 5 synthesizes the findings across all three studies, discussing what has been achieved and what should be improved. The Cumulative Thinking Prompt established that linear, structured knowledge accumulation enhances definition quality; the AUDIT Prompt demonstrated that iterative reasoning with recursive refinement yields superior and more stable performance; and the Professor-Rank-Aware Prompt revealed that fine-grained persona specifications fundamentally

shape model navigation of the reasoning space. The chapter acknowledges key limitations, including the computational cost of iterative refinement cycles, the need for adaptive termination strategies, the focus on formal mathematical quality metrics that may not fully capture pedagogical clarity for diverse learner populations, and the Western-centric academic ranks used in the persona study. These limitations point to specific areas requiring methodological refinement in future work.

Chapter 6 explores future directions, envisioning the synergistic integration and dynamic selection of the three research methodologies. Expected advances include hybrid frameworks that combine linear and iterative reasoning with context-aware persona selection, adapting strategies based on task complexity and real-time quality assessment. Another direction involves multi-agent collaboration where distinct personas with specialized reasoning behaviors work together to navigate the reasoning space. The professor-rank framework could be extended to diverse learner populations. Finally, the systematic methodology developed across the studies provides a template for optimizing LLM behavior in other domains requiring precise, structured outputs, such as scientific writing, code generation, legal drafting, and medical report synthesis.

In summary, this dissertation makes three interconnected social contributions to prompt engineering for knowledge-intensive tasks. It demonstrates that linear reasoning with structured pre-generation knowledge accumulation (Cumulative Thinking Prompt), iterative reasoning with recursive quality audits (AUDIT Prompt), and fine-grained persona specifications (Professor-Rank-Aware Prompt) each offer distinct, empirically validated advantages for mathematical definition generation. Collectively, the findings establish that deliberate prompt design, specifically, controlling the trajectory of reasoning and the precision of context, provides a reliable, training-free pathway to elicit high-quality outputs from large language models, with significant implications for educational technology and automated content generation, thus introducing positive social impacts

（様式９号）

学位論文審査の結果及び最終試験の結果報告書

山口大学大学院創成科学研究科

氏　　名	ZHU MENGYAO
審査委員	主　　査：NGUYEN PHUC HUU
	副　　査：稲葉　和也
	副　　査：福代　和宏
	副　　査：石野　洋子
	副　　査：大島　直樹
論文題目	Prompt-Guided Navigation in the Reasoning Space of Large Language Models （大規模言語モデルの推論空間におけるプロンプト誘導型ナビゲーション）

【論文審査の結果及び最終試験の結果】

本研究は、大学水準の数学用語の定義文生成を対象とし、大規模言語モデルの推論空間における探索をプロンプトによって方向づけることで、高品質な出力を得る方法を体系的に究明したものである。この課題は、形式的な厳密性と概念的な完全性とを同時に要求し、人工知能にとって認知的な負荷が大きい。近年の言語モデルは膨大な知識を保持するものの、その出力は指示の与え方によって大きく変動する。質の改善は追加学習に依拠する場合が多く、相応の計算資源を要する。

本研究の理論的中核をなすのは、出力の質を、モデルが何を知っているかではなく、その知識のどの領域へ導かれるかの問題として捉える推論空間の見方である。申請者は、この空間を部分空間とそこへ向かう方向とに分節して形式的に記述し、プロンプトを命令ではなく、その内部を航行させる誘導信号とみなす枠組みを提示している。この見方に立てば、出力の質の違いはモデルの能力ではなく、到達した領域の差として説明される。申請者はこの視座のもとで、推論の軌道と文脈指定の精度という二つの制御軸を設定し、追加学習によらずに操作しうると位置付けている。

第一の研究では、定義を直ちに書かせるのではなく、与えられた用語について、本質的なものから抽象的なものへと段階的に想起を重ねさせ、そのうえで生成に移らせる線形推論の手法を提案している。多数の数学分野にわたる用語集合を用いて既存の手法と比較し、提案手法が一貫して優位に立つことを確認している。申請者は、生成に先立つ知識の蓄積が認知的な足場となり、モデルの探索をより適切な部分空間へ収斂させると結論付けている。

第二の研究では、出力が一度きりで完結する点を問題とし、生成、品質監査、改訂の循環をプロンプトの内部に組み込む反復推論の手法を提案している。注目すべきは、監査を伴わない変種が第一の手法と同等の水準にとどまった一方、循環を備えた変種のみが性能を大き

（様式9号）

く高め、ばらつきの小ささでも際立った点である。これは、改善が知識の蓄積ではなく軌道の修正によることを意味する。申請者は、明示的な自己点検の機構が推論の軌道を動的に調整する余地を生むと論じている。

第三の研究では、文脈指定の精度そのものを検討し、教授職の階梯に沿って細分化したペルソナ設定を比較している。その結果、職位の高低と生成の質との間に明確な階層関係を見出し、広く用いられてきた優れた教授という粗い設定が最良ではないことを明らかにしている。申請者は、こうした設定を装飾ではなく、モデルを有用な情報へ導く信号として捉えている。これらの検証にあたり、申請者は大規模な用語集合と多指標による評価枠組みとを自ら整備し、人手の評価との照合により信頼性を確かめている。以上のように本研究は、推論空間を数学的に表現し、推論の軌道と文脈指定の精度を制御するプロンプト設計のみで高品質な出力を引き出しうることを実証している。精度と構造性を要する知識集約的な課題への展開が期待される方法論を示した点に意義がある。

本審査会では、予備審査会において指摘された評価結果の信頼性に関する統計的検定、および教育をはじめとする多分野への波及に関する事項について、明確な回答があった。

公聴会における主な質問内容は、評価枠組みの統計的な仕組みに関するもの、提案手法の実演に関するもの、プロンプトの違いによる出力内容の差異などに関するものであった。いずれの質問に対しても発表者から的確な回答がなされた。

以上より本研究は独創性、信頼性、有効性、実用性ともに優れ、博士（学術）の論文に十分値するものと判断した。

論文内容及び審査会、公聴会での質問に対する応答などから、最終試験は合格とした。

なお、主要な関連論文の発表状況は下記のとおりである。（関連論文　計3編、参考論文　計0編）

(a) 査読のある雑誌等

1) Mengyao Zhu and Phuc Huu Nguyen, A Survey of LoRA Algorithm Variations for Language Models, Lecture Notes in Computer Science (LNCS), Springer, Vol.15836, pp.275-290, DOI: 10.1007/978-3-031-97141-9_19, 2025.
2) Mengyao Zhu and Phuc Huu Nguyen, A Survey of Algorithmic and Contextual Decomposition Methods Across Language Model Pipelines, Lecture Notes in Artificial Intelligence (LNAI), Springer, Vol.16354, pp.55-67, DOI: 10.1007/978-981-95-4960-3_5, 2025.
3) Mengyao Zhu and Phuc Huu Nguyen, Reasoning-Aware Instruction-Tuned Language Models are Competitive Sentiment Classifiers, Lecture Notes in Networks and Systems (LNNS), Springer, Vol.1904, pp.307-320, DOI: 10.1007/978-3-032-20897-2_23, 2026.