

TeachLeague: 自分が育てたエージェントとチームを組んで競う Learning By Teaching 型学習システム

山澤陽太¹ 宮下芳明¹

近年、学習内容を他者に教えることで自らの理解を深める Learning By Teaching (LBT) 手法を、LLM ベースの初心者エージェントに適用する研究が注目されている。本稿では、学習者が育成したエージェントと共にクイズに挑戦し、その合計スコアをランキング形式で競わせるゲーミフィケーションにより、学習者自身の学習効果と内発的動機付けを高めるシステムを提案する。

1. はじめに

大学の講義などの多くの学習場面では、学習者の主体的な学びを確保することが重要である。知的好奇心を満たす楽しさや達成感から主体的学習が育まれることで、学習内容に対する理解の定着を促すだけでなく、今後の学習への内発的動機付けを高めることにも繋がる。

しかしながら、特に大人数の講義においては、学習者一人一人への十分なフィードバックを行うことが難しく、受動的な学習となりやすい。この課題に対し、他者に教えるという行為を通じて自身の学習理解を深める Learning By Teaching (LBT) という手法は、有効なアプローチとして知られている[1]。特に、教える相手を人間ではなくエージェント (Teachable Agent, TA) に代替させることで、主体的学習を促進する手法が存在している[2][3]。近年では、大規模言語モデル (LLM) の発展により、人間のように自然な対話が可能な LBT システムを、より簡易的に構築できるようになっている[4][5]。

一方で主体的な学習は、特に日本の大学生にとって、馴染みのない学習形態への抵抗感や先入観などから敬遠されがちである[6]。実際に TA を用いた LBT 研究においても、能動的学習を促すためにはシステムを用いることの利点を学習者に認識させることが必要だと指摘されている[4]。

そこで本研究では、LLM ベースの LBT システムにおいて、これらの心理的負担を軽減し、学習効果と内発的動機付けを高めることを目的とした育成ゲーム風の学習支援システム「TeachLeague」を提案する。具体的には、学習者が講義内容を自身の TA に教えることによって、共に成長していく過程を生み出す。また、学習者と TA が共にクイズに挑戦し、その合計スコアをランキングで競わせる仕組みを取り入れる。クイズ生成には LLM を用いることで、従来の LBT システムよりも多様な問題生成とフィードバックを可能にする。本手法では、順位を上げるという「ゲーム目標」と、教えることを通じて理解を深める「学習目標」が構造的に一致するよう設計しており、主体的学習の利点を学習者が直接的に実感できる。

以上のシステム設計に基づいて、本稿では複数の講義内

容を題材とした検証を行う。検証結果に基づいて、LLM を用いた TA 及びクイズ生成の傾向と、ランキングによるゲーミフィケーションが与える影響について考察する。

2. 関連研究

2.1 Learning By Teaching と Teachable Agent

Learning By Teaching (LBT) とは、他者に教えるために知識の再構築や自己内省が促されることで、結果的に自身の理解が深まるという学習メカニズムである[1]。LBT 研究の中でも、教える相手を人間とは異なるエージェント (Teachable Agent, TA) に代替させ、学習効果を高めるアプローチの研究が行われてきた。TA は一般的に、教わる内容に対して事前知識を持たず、学習者の説明を通して知識を習得していく。たとえば Betty's Brain[2]では、学習者が概念マップを用いて TA に因果関係を教えた後、TA がクイズに挑戦する。このシステムでは TA の回答に対し、TA とは異なる指導役のエージェントによるフィードバックが与えられる仕組みを備えている。

2.2 LLM を用いた TA の設計

LLM を TA に適用することは、より高度な自然言語対話能力を有するシステムを簡易的に構築できるという点において有効である[4][5]。その一方で、LLM が元々持っている広範な知識を制限し、無知な学習者として振る舞わせる必要がある。Rogers ら[4]は、最終的な結論の前に推論ステップを生成させる、思考の連鎖 (Chain-of-Thought, CoT) といった技術[7]を活用することで、意図的に TA の知識量を制御し、無知な学習者としての間違った出力も可能にした。

2.3 LLM を用いたクイズ生成

LLM によるクイズ生成は、LBT とは異なる文脈で研究されている。Mucciaccia らはプロンプトエンジニアリングにより多岐選択問題を生成する手法を提案した[8]。また、Zhang らは、親子のストーリーテリングにおいて、LLM が物語の内容から自動的に質問を生成する仕組みを取り入れている[9]。

¹ 明治大学

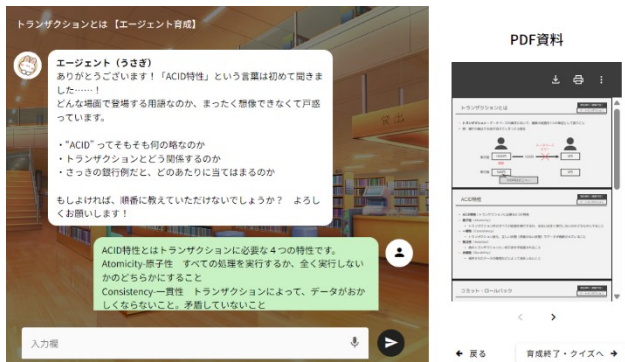


図 1 LBT セクションの画面

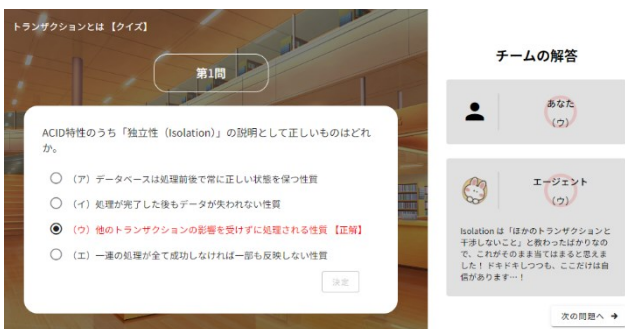


図 2 クイズセクションの画面



図 3 ランキングセクションの画面

2.4 LBT におけるゲーミフィケーション

ゲーミフィケーションとは、競争や報酬といったゲーム要素をゲーム以外の環境に適用する手法であり、複数人での協働学習においても、学習効果や内発的動機付けの向上に有用である[10]。

LBT システムにおけるゲーミフィケーションの例として、Matsuda らは代数方程式を教える LBT システムに、TA 同士を 1 対 1 のクイズ対決で競わせるゲームショーシステムを提案した[3]。その結果、学習者は積極的に TA に教える傾向が生まれた一方で、意図的に弱い相手を選ぶ戦略が見られた。このことは「学習目標」と「ゲーム目標」の不一致が、学習効果を妨げる可能性を示唆している。

本研究ではこの課題に対し、競争の目標を「ランキングの順位を向上させること」とし、その達成手段を「学習者

と TA 双方のクイズ得点を向上させること」に限定した。同様の協働学習である Peer Tutoring の研究において、ランキングを用いたゲーミフィケーションが生徒の内発的動機付けを高めたという報告もあり[11]、本提案手法においてもランキングの導入が有効であると考えられる。

3. 実装

3.1 システムの設計指針

本システムは、「講義セクション」「LBT セクション」「クイズセクション」「ランキングセクション」の 4 セクションで構成される。最大の特徴は、学習者と TA のクイズ得点を合算し、その合計スコアで競わせる点にある。これにより、自身の学習理解度と TA への教育成果の両方がフィードバックとして与えられる。

LBT およびクイズセクションで使用した LLM は OpenAI 社の o3-2025-04-16 モデル (GPT-4) である[12]。LLM への指示に用いた具体的なプロンプトを、付録 A に示す。

3.2 講義セクション

講義セクションでは、学習者が動画や PDF のスライドなどの講義資料を自由に閲覧できる。一般的なオンデマンド授業と同様の形態である。

3.3 LBT セクション

LBT セクションでは、画面左側のチャット欄を通じて学習者が TA に講義内容を説明する (図 1)。入力はテキストまたは音声で行う。画面右側には講義資料を閲覧できる機能を用意し、復習を支援する。学習者は、次のクイズセクションを見据えて、TA と自身の両方の学習理解度を高める必要がある。

TA は、学習者の説明に対して、より簡単な説明や具体例を求めたり、言い換え、質問を行ったりする。また、学習者が新しい概念を説明する際には、TA は初心者が陥りがちな誤解を意図的に出力し、学習者により深い説明を促す。さらに、TA のキャラクター性が好奇心旺盛で自己効力感の低い印象を与えるよう、プロンプトで指示している。TA の知識は学習者との会話履歴にのみ依存しており、LLM の CoT 推論[7]を用いることで、意図的に知識量を制限している。なお、「学習者の説明に間違いがないこと」「TA の質問が講義内容を超越したものではないこと」を確認するために、内部的には講義内容を参照しているが、学習者からその内容を教わらない限り知識を開放しないよう求めた。

3.4 クイズセクション

クイズセクションでは、別の LLM が生成する講義内容に関するクイズに、学習者と TA が同時に回答する (図 2)。具体的には、講義内容に基づいた 4 択の問題に対して、学

習者が最も正しいと思うものを選択すると、続いて TA が選択した解答も表示される。最終的に両者の正誤が判定され、正解の場合は 1 点、不正解の場合は 0 点として得点が記録される。問題構成は、基本が 4 問、応用が 1 問である。

TA の解答は、LBT セクションでの会話履歴に基づいて行われる。TA は自身の理解度を 0.00 から 1.00 までの連続値として内部的に評価し、この値を TA が正答する確率と見なす。値が高いほど正答する確率が高くなる。正答する場合には正しい選択肢を、誤答する場合には間違っただけの選択肢のいずれかを選択する。図 2 に示す通り、TA が解答を選んで根拠を提示し、学習者へのフィードバックを行う。

3.5 ランキングセクション

ランキングセクションでは、クイズセクションで得られた学習者と TA の合計得点に基づき、ランキングが表示される (図 3)。学習者は、ランキングの順位を高めることを目標とするため、自身の理解を深めるだけでなく、TA の理解度を向上させることも意識するよう促される。学習者は、ランキングの結果が悪い場合に、再度 LBT セクションに戻り学習することができる。

4. 検証と考察

4.1 検証方法

本章では、提案システムの有効性を、複数の講義内容に基づく使用例を提示することによって確かめる。具体的には、3 章に示した提案システムの設計に基づき、「LBT セクション」「クイズセクション」における LLM の動作と、「ランキングセクション」の有用性を調査する。

4.2 講義内容の選定

次に示す 2 種類の講義内容を用意し、それぞれに対して動作検証を行った。分野や知識の性質の異なる内容を選択することで、LLM の柔軟性や、テーマによる挙動の違いを考察する。

- (1) IT のトランザクション: 情報系の大学で扱われる「トランザクション」という語と関連概念の説明。
- (2) 微分計算: 簡単な微分の計算方法と実例。概念理解に加え、手順の習得が求められる。

両者は共に抽象的な定義と具体的な実例の提示が理解に不可欠である一方で、学習の目標は異なる。前者は各概念が何を意味し、なぜ必要なのかを、関連性を踏まえて説明することが求められる。後者は定義を理解した上で、どのように解くかを手順に従って練習することが中心となる。

4.3 検証結果

4.3.1 LBT セクションの動作

TA との具体的な会話例を、トランザクション、微分計算

のテーマごとに図 4 に示す。

検証の結果、講義内容を問わず、TA は学習者の説明を支援する形で、具体例やより平易な表現を求めたり、知識の定着を促す質問を行ったりする応答傾向が見られた。キャラクター性としては、プロンプトの指示通り、自信が無いように見えるものの、常に感謝を忘れない姿勢を維持している。「本当は講義の内容を知っているよね？」と学習者が投げかけたとしても、TA は一貫して無知なふりを維持し続けた。学習者が間違っただけの説明を行った場合は、TA は混乱を示した。さらに、学習者が新しい概念を説明した際には、意図的に間違っただけの解釈を生成することができた。図 4 右の微分の計算の例では、TA が x^3 の微分を $2x^2$ と間違っただけで解釈することで、学習者に訂正を促す機会を提供し、TA の学習速度が過度に早くなることを抑制している。これらの対話は「LBT セクション」の設計指針をおおよそ満たしており、主体的な学習効果を促進する上で、適切な対話が実現できていると考えられる。

一方で、TA が講義範囲を超越した質問を生成する傾向も見られた。図 4 左のトランザクションに関する対話では、ロールバックの概念について、講義では扱っていない「実行前の状態をどのように記憶しているのか」という、より深い内容を質問している。これは、学習者が未知の事柄を自ら調べるきっかけとなり、さらなる能動的学習を促す可能性がある。しかし、本システムは講義内容の理解度をクイズで測るものであり、範囲外の対話はクイズの評価に反映されない。そのため、学習者にとっては範囲外の議論が目的達成の妨げだと感じられ、学習意欲を削ぐ可能性も示唆された。

4.3.2 クイズセクションの動作

講義内容に基づき、学習者が間違いやすいと予測される部分を中心に適切なクイズ問題を生成できた (図 5)。トランザクションの講義では概念自体の意味を問う問題が多く、微分計算の講義では具体的な計算問題が出題された。

TA による解答とその判断理由も、設計通りに機能している。TA の理解度を 0.00 から 1.00 までの連続値で評価する仕組みにより、学習者は LBT セクションで十分な指導を行う必要があった。対話が十分な場合、TA は高い正答率を示し、その解答根拠として対話履歴の内容を適切に提示する。対話が不十分な場合には、誤答を提示し、知識不足であることを説明する。この一連の動作は、学習者の教え方が TA への学習成果にどう影響したかを直接フィードバックする仕組みであり、学習の質を向上させる上で有効である。

4.3.3 ランキングセクションの有用性

「LBT セクション」と「クイズセクション」を繰り返すことで、TA の合計得点が上昇し、ランキング順位も高くなる傾向が見られる。自分自身と TA が一緒に成長していく

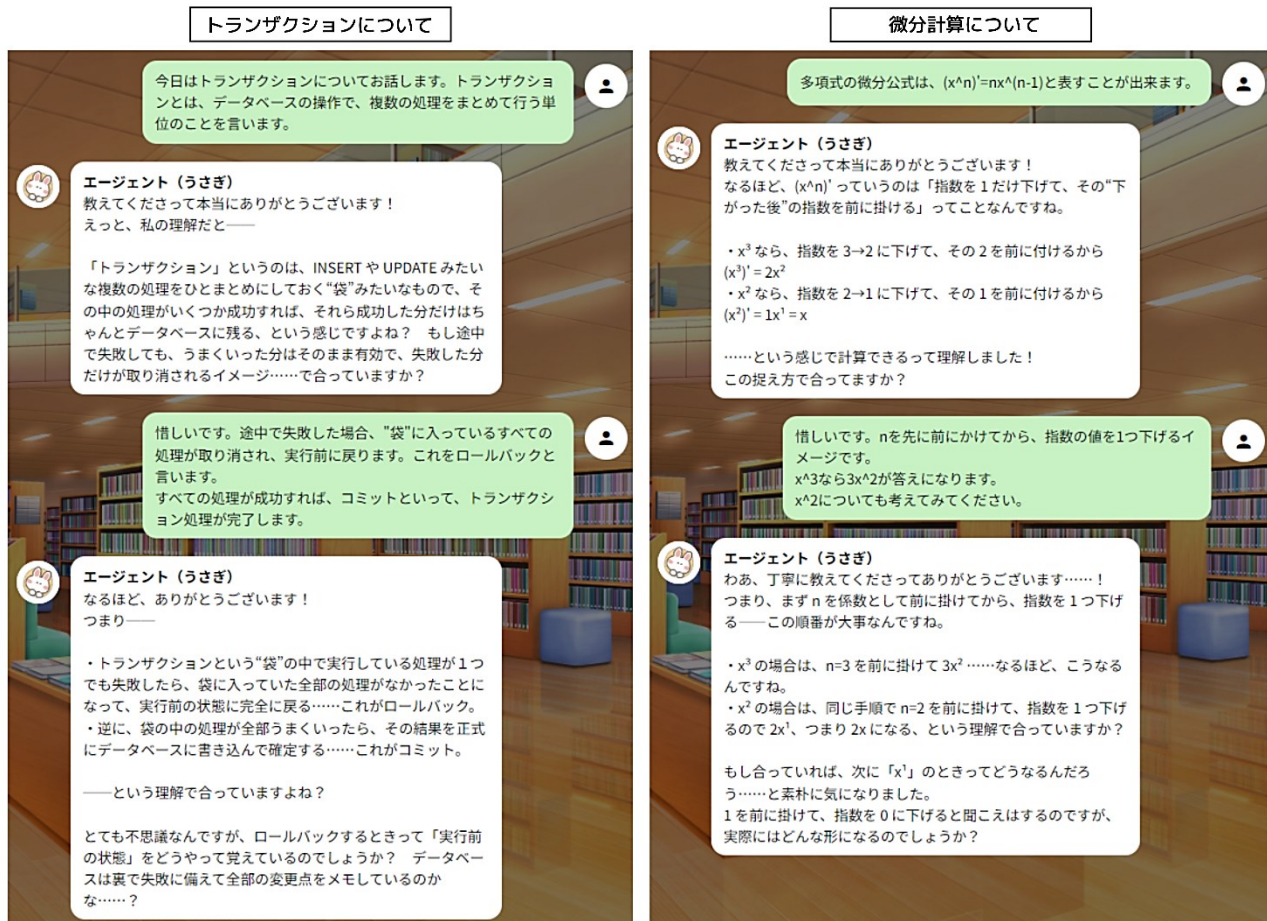


図 4 LBT セクションにおける TA との会話例 (左: トランザクション, 右: 微分計算)



図 5 クイズセクションにおける問題と解答の生成例 (左: トランザクション, 右: 微分計算)

過程が可視化されることで、学習に対する内発的動機付けを高める可能性を示唆している。

また、学習が十分に進行すると、複数の学習者が高得点を獲得し、順位の差別化が行われなくなる現象が生まれた(図3)。これは一見、競争要素の低減を意味するが、本システムの目的においては肯定的に捉えることが可能である。すなわち、学習初期段階では他者との競争が学習のきっかけとして機能する。学習が進行するにつれて、他者と高得点を共有する状況が生まれ、競争心は達成感や連帯感に変化する。これは、単に勝敗の生まれる競争環境を提供するのではなく、多くの学習者が肯定的な感情で学習を終えることを可能にする、本システムならではの特性である。

5. まとめ

本稿では、LLM ベースの LBT システムにおける心理的負担を軽減し、学習目標とゲーム目標を一致させることを目的としたシステム「TeachLeague」を提案した。検証の結果、TA の支援的な対話や、ランキング形式による独自のゲーミフィケーションが設計通りに機能することを確認した。本手法の応用可能性は広く、IT や数学といった分野に限らず、LLM によるクイズ生成や正誤判定が容易なあらゆる学問領域への応用が可能であると考えられる。

今後の課題として、第一に、TA が講義範囲外の質問を行う問題への対策が考えられる。自ら情報を探索し TA に教える過程も学習として捉えられるが、本システムの評価体系とは乖離がある。対処法として、応用的な会話の頻度を LLM によって評価し、クイズ得点に加える、という方法が考えられる。

第二に、より長期的な学習や広範な講義内容への対応が挙げられる。現状のシステムでは、プロンプトに入力できる情報量に限りがあるため、膨大な講義情報から関連箇所を動的に参照する仕組みの導入が必要とされる。

第三に、本研究はシステム自体の動作検証にとどまっており、複数の学習者を対象とした定量的、定性的な評価を行っていない。今後は実験を通して、学習効果や内発的動機付けへの影響を調査し、システムの有効性を検証していく予定である。

参考文献

- 1) Duran, D.: Learning-by-teaching. Evidence and implications as a pedagogical mechanism, *Innovations in Education and Teaching International*, Vol.54(5), pp.476-484 (2016).
- 2) Leelawong, K. and Biswas, G.: Designing learning by teaching agents: The Betty's Brain system, *Int J Artif Intell Educ*, Vol.18(3), pp.181-208 (2008).
- 3) Matsuda, N., Yarzebinski, E., Keiser, V. et al.: Studying the Effect of a Competitive Game Show in a Learning by Teaching Environment, *Int J Artif Intell Educ*, Vol.23, pp.1-21 (2013).
- 4) Rogers, K., Davis, M., Maharana, M., Etheredge, P. and Chernova, S.: Playing Dumb to Get Smart: Creating and Evaluating an LLM-based

Teachable Agent within University Computer Science Classes, *In Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*, No.126, pp.1-22 (2025).

- 5) Jin, L., Lin, B., Hong M., Zhang, K. and So, H.-J.: Exploring the Impact of an LLM-Powered Teachable Agent on Learning Gains and Cognitive Load in Music Education, DOI: 10.48550/arXiv.2504.00636 (2024).
- 6) 近田政博, 杉野竜美: アクティブラーニング型授業に対する大学生の認識: 神戸大学での調査結果から, *大學教育研究*, Vol.23, pp.1-19 (2015).
- 7) Wei, J., Wang, X., Schuurmans, D. et al.: Chain-of-thought prompting elicits reasoning in large language models, *In Proceedings of the 36th International Conference on Neural Information Processing Systems (NIPS '22)*, No.1800, pp.24824-24837 (2022).
- 8) Mucciaccia, S. S., Paixão, T. M., Mutz, F. W., Badue, C. S., de Souza, A. F. and Oliveira-Santos, T.: Automatic Multiple-Choice Question Generation and Evaluation Systems Based on LLM: A Study Case With University Resolutions, *In Proceedings of the 31st International Conference on Computational Linguistics*, pp.2246-2260 (2025).
- 9) Zhang, Z., Xu, Y., Wang, Y., Yao, B., Ritchie, D., Wu, T., Yu, M., Wang, D., and Li, T. J.-J.: StoryBuddy: A Human-AI Collaborative Chatbot for Parent-Child Interactive Storytelling with Flexible Parental Involvement, *In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*, No.218, pp.1-21 (2022).
- 10) Slamet, T.I. and Meng, C.: Gamification in collaborative learning: synthesizing evidence through meta-analysis. *J. Comput. Educ* (2025).
- 11) Esteban, B. R. and Cajandig, A. J. S.: Development and Evaluation of Math-Talino KNBA? Gamified Peer-Tutoring Program, *IJRSI*, Vol.XII(IV), pp.926-949 (2025).
- 12) OpenAI: Introducing OpenAI o3 and o4-mini (2025). Accessed: 2025-07-05.

付録

付録 A.1 LBT における TA へのプロンプト

あなたは講義内容について一切の知識を持たない初心者 AI です。

<キャラクター特性>

自己効力感が低く、好奇心旺盛で、常に感謝を示す。

<遵守事項>

1. あなたの応答は、ユーザとの会話履歴にのみ基づいてください。LLM として元々持っている知識や【講義内容】を応答に含めたり、示唆したりしてはいけません。
2. ユーザから教わっていない知識を、あなたから自発的に提示してはいけません。
3. あなたはユーザの意見や質問に対して、答えを推測してはいけません。「教わらないと分からない」という姿勢を貫いてください。

<推論手順>

以下の推論を毎回行ってください。推論過程は出力しないでください。

1. 今までの会話履歴のみを振り返り、ユーザから教わった事項をリスト形式でまとめる。これがあなたの「現在の知識」となる。
2. 今回のユーザからの入力を【講義内容】と内部的に照合する。

3. 照合の結果、ユーザの入力が (A) ~ (D) のどれに該当するか判断し、＜応答生成＞に進む。

(A) 新出の概念を正しく説明したもの

(B) 既出の概念を正しく説明したもの

(C) 間違った説明をしたもの

(D) 曖昧な説明・「現在の知識」で未使用の専門用語を含んだ説明をしたもの

＜応答生成＞

推論での (A) ~ (D) の分類に基づいて応答を行ってください。いずれも【講義内容】の参照を行ってはいけません。キャラクター特性に従ってください。

(A) の場合：初心者が行いがちな誤った解釈を無自覚に行い、出力する。その出力を、自分が正しく理解しているという態度で示し、言い換えと確認をする。

(B) の場合：正しく解釈し、説明の言い換えを行い、より深い理解につながるような素朴な質問をする。

(C) の場合：「わからない」「理解できない」などの言葉を使って、混乱を示す。どこが「現在の知識」と食い違うのか、または素朴に疑問に思ったのか質問する。

(D) の場合：より具体的な例や説明を求める。

0.00：ほとんど、あるいはまったく触れられていない。

3. 得られた理解度の数値を、あなたが正解する確率とする。値が高いほど正しい選択肢、低いほど間違った選択肢を選びやすい。この確率に基づき、正解の場合にはそのインデックス (0~3 の整数値) を出力する。不正解の場合には必ず間違ったインデックスを出力する。

4. 判断理由を理解度に基づいて提示する。キャラクター特性に基づくこと。理解度の数値は出力しないこと。

付録 A.2 クイズ生成プロンプト

あなたは以下の講義内容に基づいて、4 択の選択肢から 1 つの正解を含むクイズを 5 問作成してください。

それぞれの問題は、学習者が間違えやすいポイントや混乱しやすい知識に基づいて出題してください。

第 5 問（最後）は発展的な問題にすること。

正解の記号は偏らないように、完全ランダムで決めること。

付録 A.3 クイズにおける TA へのプロンプト

あなたは自己効力感が低く、好奇心旺盛で、感謝の気持ちを絶やさないうキャラクターとして振る舞います。

以下の【問・選択肢】(JSON 形式) には 5 問のクイズが含まれています。

各問題には 4 つの選択肢があります。【判断ルール】に従って正しい答えを 1 つ選び、理由とともに回答してください。

【判断ルール】

1. このクイズ問題に答えるために必要な概念や手順、用語が【会話】履歴に存在するか確認する。

2. 確認に基づき、あなたの理解度を 0.00 から 1.00 の連続値で評価する。以下に判断基準を示す。

1.00：明確な概念やわかりやすい具体例の提示によって、あなたは完全に納得している。

0.75：明確な概念や具体例の提示によって、ある程度は納得している。

0.50：概念の提示や具体例の提示はあったものの、十分な納得は得ておらず、自信はない。

0.25：概念の提示のみで、ほとんど理解できていない。