

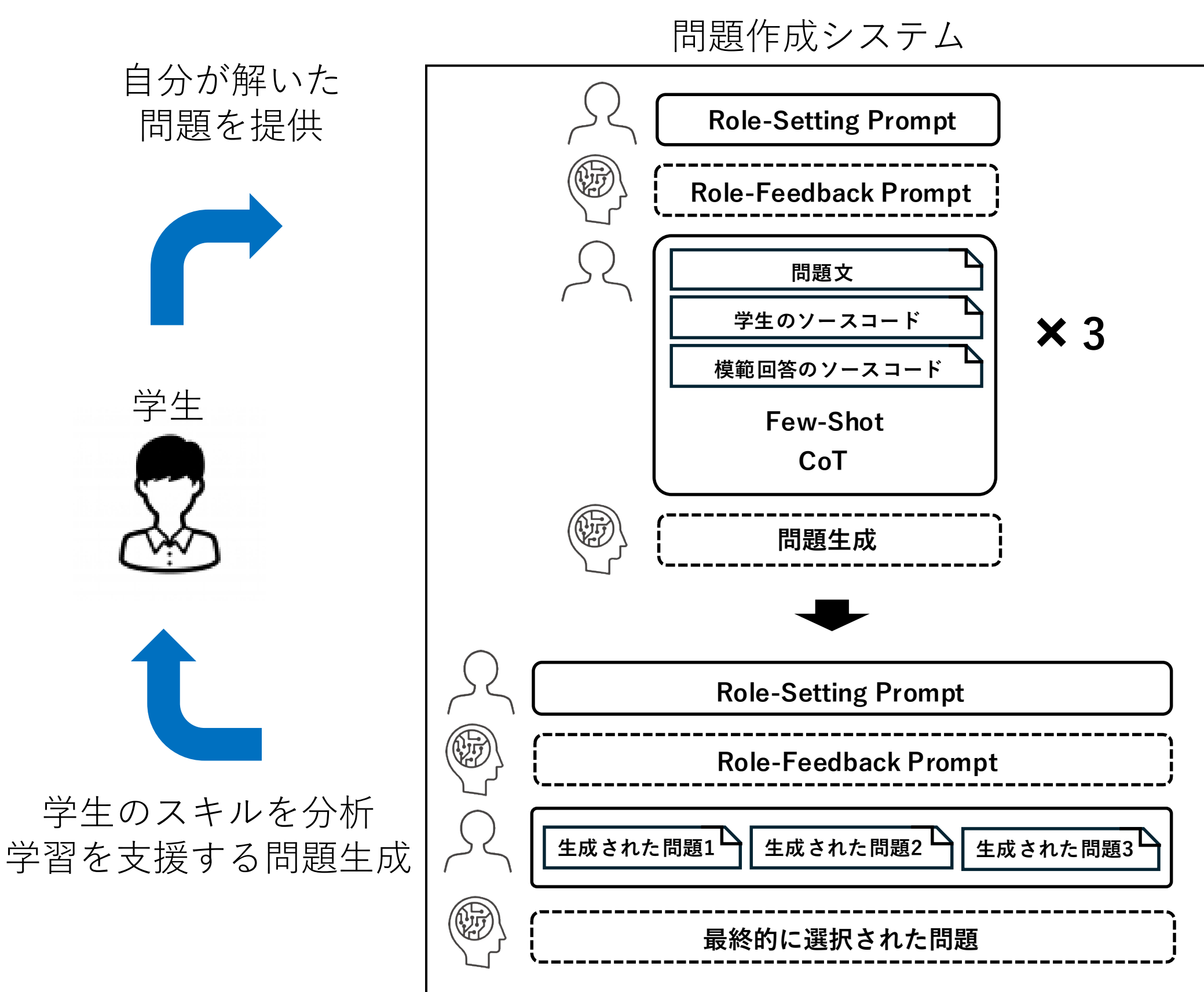
LLMによるプログラミング初学者の スキル分析と問題作成による学習支援

長谷川 駿一
(制御・情報システム工学専攻)

大枝 真一
(情報工学科)

研究背景・目的

プログラミング初学者の能力判定と学習支援について、本研究では、LLMの持つ自然言語処理能力を最大限に活用し、学習者のスキル分析と個別最適化された問題生成を同時に行うことを目指す。具体的にはプロンプトエンジニアリングの手法を組み合わせることで、LLMに教育者としての役割を与え、学習者のソースコードを分析させ、その分析結果に基づいて個々の学習者に最適な問題を自動生成するシステムの構築を試みる。これは、従来の機械学習アプローチでは困難であった、学習者の問題解決プロセスに対する質的評価と、それに基づく個別最適化された学習支援の実現という、教育工学上の重要課題に対する新たな解決アプローチを提示することを目的とする。



プロンプトエンジニアリング手法

• Role-Play Prompting

LLMに特定の役割を付与し、その役割に沿った応答を生成させることで、より高度な推論を引き出す。

• Chain-of-Thought (CoT)

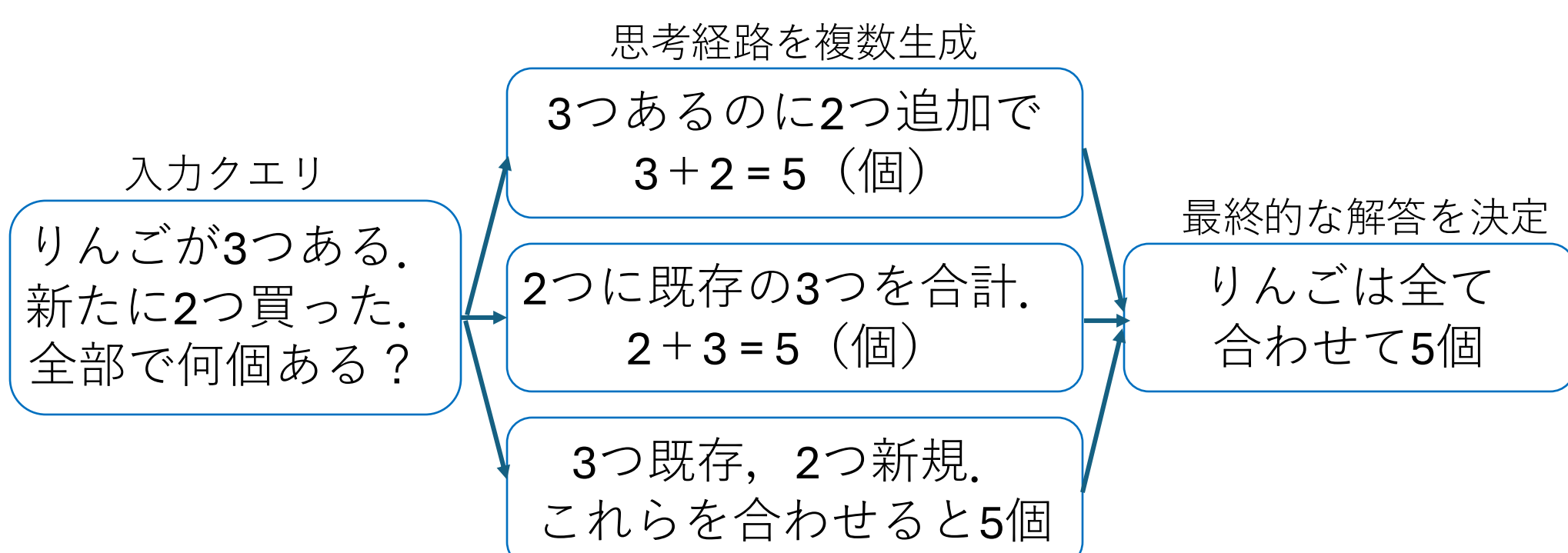
回答に至るまでの思考過程の例をプロンプトに含めることで、LLMに段階的な思考過程を示すよう促す。

• Few-Shot Learning

モデルに対する入力クエリ内に、説明文と少数の例示を包含することでLLMにタスクの本質を理解させる。

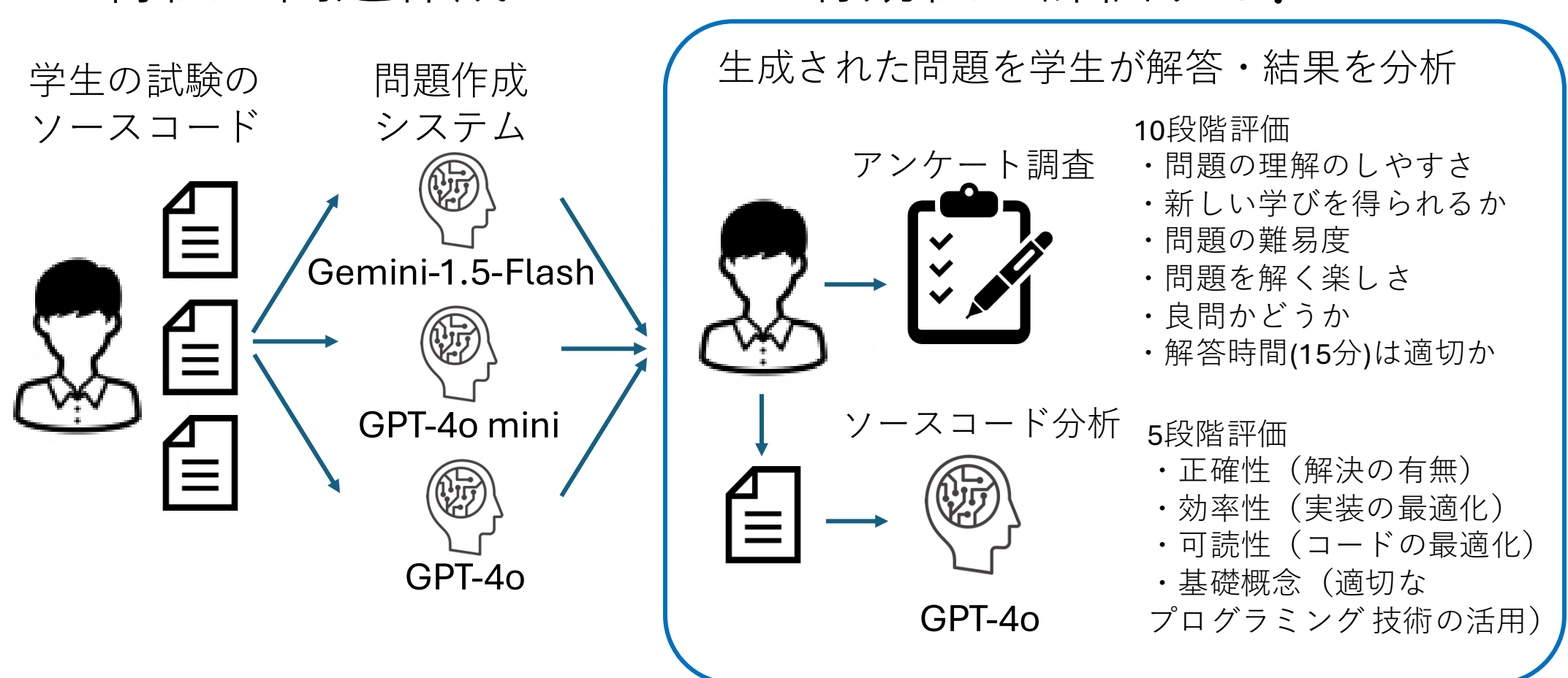
• Self-Consistency

複数の異なる推論経路をサンプリングし、それらの中で最も一貫性のある答えを選択する。



実験概要

2024年度木更津工業高等専門学校情報工学科2年の「プログラミング基礎II」後期中間試験に出席した学生41名が提出したソースコードを用いて3種類のLLM (Gemini 1.5 Flash, GPT-4o mini, GPT-4o) を使用した問題生成システムで問題を生成させた。生成した問題を解いた感想のアンケート結果と解いた際に書いたソースコードをLLM(GPT-4o)に分析させ各LLMの特性と問題作成システムの有効性を評価する。



実験結果

1. アンケート結果の平均の比較

各アンケート項目に対して一元配置分散分析を実施
→ 問題の「理解しやすさ」に有意差を確認

「理解しやすさ」に対して TukeyのHSD検定を実施
→ Gemini 1.5 Flash と GPT-4o mini の間に有意差を確認

アンケート項目の一元配置分散分析

| 評価項目 | F値 | p値 | 有意性 (5%) |
|--------|-------|-------|----------|
| 理解しやすさ | 3.093 | 0.050 | あり |
| 新しい学び | 0.390 | 0.678 | なし |
| 難易度 | 1.503 | 0.227 | なし |
| 楽しさ | 0.532 | 0.589 | なし |
| 良問評価 | 0.297 | 0.744 | なし |
| 時間の適切さ | 0.100 | 0.905 | なし |

「理解しやすさ」のTukeyのHSD検定

| モデル比較 | 平均差 | p値 | 有意性 (5%) |
|---------------------------------|--------|--------|----------|
| GPT-4o GPT-4o mini | -0.972 | 0.214 | なし |
| GPT-4o Gemini 1.5 Flash | 0.428 | -0.969 | なし |
| GPT-4o mini Gemini 1.5 Flash | 1.400 | 0.046 | あり |

- Gemini 1.5 Flashは問題文章の組み立てに優れる。
- GPT-4o miniは「楽しさ」「良問評価」に秀でており学生のモチベーションを高める問題生成に優れる。
- GPT-4oは上記2モデルの間の特徴を持つ。

2. アンケート項目とソースコード評価結果の相関

- 特にGemini 1.5 Flashで「難易度」が高いと評価されるとコードの品質が大きく低下する。
- Gemini 1.5 Flashで「新しい学び」が大きいほどコードの品質が大きく低下する。
- GPT-4o miniやGPT-4oにはそのような傾向がない。
→ **Gemini 1.5 Flashは一貫した問題生成能力を持つ。**

まとめ

LLMを用いた問題生成は、学習意欲向上に寄与する可能性を示唆する。しかし、難易度調整、時間配分、コード品質に課題が残る。LLMには、学習履歴に基づく適応的な難易度調整、最適な解答時間の提示、生成コードに対する詳細なフィードバック機能の実装が求められる。