

生成 AI の教育利用に向けた正確性と学習者への影響の検証 ープログラミング学習アプリの開発と評価ー

Verification of Accuracy and Impact on Learners for Educational Use of Generative AI - Development and Evaluation of a Programming Learning Application -

山本 泰輝^{*1}, 宮原 一弘^{*1}
Taiki YAMAMOTO^{*1}, Kazuhiro MIYAHARA^{*1}

^{*1}名古屋市立大学大学院理学研究科
^{*1}Graduate School of Science, Nagoya City University
Email: c237720@ed.nagoya-cu.ac.jp

あらまし：本研究では、学校現場における生成 AI の利用に関する議論に寄与することを目的とし、生成 AI の正確性や学習者が受ける影響を検証した。また、この検証を行うために生成 AI を用いた学習用アプリケーションの開発を行い、この学習アプリを使用した実験を行った。その結果、一定の正確性が認められたが教育利用には不十分な水準であり、さらなる生成 AI の進化が必要であることが示唆された。また、生成 AI に依存する例も見受けられ、教育利用には学習者側への指導が必要であることが示唆された。

キーワード：生成 AI, Society5.0, 個別最適な学び, プログラミング教育

1. はじめに

近年、人工知能（AI: Artificial Intelligence）の発展が著しく、特に生成 AI（Generative AI）はユーザーが入力したプロンプト（指示文）やデータから新たなコンテンツを作り出すことができるため、多くの分野で活躍が期待されている。これは学校現場においても例外ではなく、教員の労働実態の改善や個別最適化された教育の実現など教育現場の抱える課題を解決する可能性がある。

実際、中央教育審議会においても生成 AI は「業務の効率化や質の向上など、働き方改革の一環として活用を推進する必要がある。」⁽¹⁾とされている。しかし、生成 AI の教育利用に多くの課題が残されており、文部科学省の「初等中等教育段階における生成 AI の利用に関する暫定的なガイドライン」⁽²⁾では、ハルシネーションを始めとした生成内容の正確性に関する問題や生成 AI の依存による学習者の自律性の低下の可能性などが指摘されている。そのため、さらなる生成 AI の知見の蓄積や成果・課題の検証が必要であると考えられている。

そこで本研究では、学校現場における生成 AI の利用に関する議論に寄与することを目的とし、生成 AI の正確性や学習者が受ける影響を検証した。また、この検証を行うために生成 AI を用いた学習用アプリケーション（以下、学習アプリ）の開発を行い、この学習アプリを使用した実験を行った。

2. 学習アプリの開発

今回開発した学習アプリは、問題の生成、質疑応答、正誤判定の各機能を生成 AI が担う設計とした。いずれの機能でも、システムは学習者（ユーザー）の入力を受け取り、それを基にプロンプトを作成して生成 AI に送信する。生成 AI の出力結果はシス

テム内で加工され、学習者にとって分かりやすい形式で画面上に表示される。プロンプトは、事前に設定された学習対象の言語や生成 AI の役割を基に、学習者からの入力内容と組み合わせて生成される。また、生成 AI との通信には API（Application Programming Interface）を使用する。

開発した学習アプリの画面を図 1 に示す。学習アプリは表示される画面を HTML と CSS によって、Web サーバー上で起動する Web アプリを JavaScript によって実装した。また、Web サーバー上に MySQL によるデータベースを構築することで、学習者の入力や生成 AI の生成文などの実験に必要な情報を記録することを可能にした。



図 1 学習アプリの画面

3. 実験

名古屋市立大学総合生命理学部および同大学院理学研究科の学生を始めとした計 19 名を被験者として実験を行った。被験者のプログラミング経験や生成 AI の利用経験の有無は各々異なる。

実験の流れは次の通りである。

- ① 学習アプリの利用
- ② 「使用感に関するアンケート」への解答
- ③ ①,②を3回繰り返す
- ④ 「全体アンケート」への解答

なお、今回の実験では被験者の負担も考慮し実験の所要時間を1時間としたうえで、学習アプリの利用を一人当たり3回行ってもらった。

4. 結果と考察

4.1 生成 AI の正確性に関する検証

MySQL 上に記録された被験者が解答したソースコードを生成 AI を用いることなく、研究室の PC 上でコンパイルおよび実行した。その実行結果と生成 AI の判定結果を比較し、表1の混同行列を作成した。

表 1 混同行列

		生成 AI の判定結果		合計
		正解	不正解	
実際の 実行結果	正解	49	6	55
	不正解	11	96	107
合計		60	102	162

表1より、より正解率は0.89506、適合率は0.81667、再現率は0.89091、F1スコアは0.85217となった。約89%のコードに対して正しく判定結果を出しているものの、実際の学習利用としては不十分と考えられる。特に、適合率が約81%と低く、偽陽性（誤って正解と判定されたもの）のコードが一定数存在していたことを示している。

偽陽性となったコードと講評文を分析すると、出力結果が題意に沿っていないにも関わらず正解と判定されたケースと、アルゴリズムの欠陥やヘッダーファイルを読み込むための記述の欠落といったプログラムの実行に影響する重大なミスを含んでいるにも関わらず正解と判定されたケースの2つのパターンが見られた。前者の要因は、同様のミスを不正解と判定する場合もあり、また、講評文中にミスを指摘する文言が見られることから、生成 AI の判定基準に揺らぎがあったためだと考えられる。そのため、プロンプト内で判定基準を明確に記述することで改善できる可能性がある。一方、後者は生成 AI がコードの構造や意味を十分に理解できていないことを示唆しており、生成 AI の学習量が足りないことが原因であると考えられる。

4.2 生成 AI が学習者に与える影響に関する検証

次にデータサーバー上に記録された利用データとアンケートの結果を組み合わせることで被験者がどのように学習アプリを利用していたのかについて分析した。

質疑応答機能は計54回のアプリ利用のうち41回

で使用され、合計で110回の質疑応答が行われた。1回のアプリ利用あたり約2.97回であり、最小値は1回、最大値は13回であった。質疑応答の使用回数に関するグラフを図2に示す。

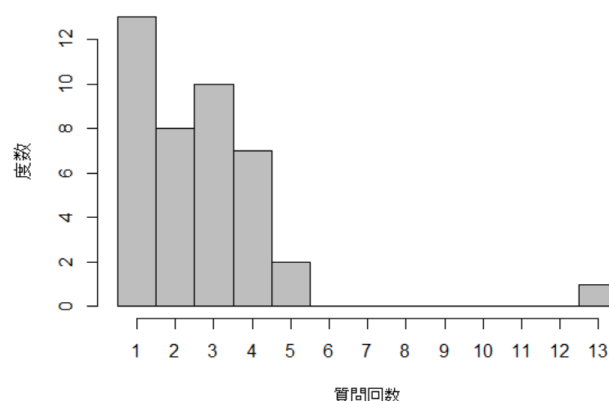


図 2 質問機能の使用回数

多くの被験者が1～5回と質疑応答を適切に利用しているのに対し、13回と質疑応答を多用する例も見られた。また、質疑の内容を確認したところ、問題文をそのまま複製・貼り付けしたものもあり、その数は20回(全質問の約18.2%)であった。

以上のことより、生成 AI への依存や回答の流用などにより学習の機会を奪ってしまう可能性があることが分かった。これらを防ぐため、質疑応答の回数に制限をかけることや、質疑応答のプロンプト内に問題文を盛り込んだうえ問題に直接回答しないようにすることが必要であると考えられる。

5. 結論

本研究では、文部科学省のガイドラインに挙げられた懸念される課題を完全に払拭する結果を得ることはできなかった。現状の生成 AI は課題が多く、さらなる改善が必要であると考えられる。しかし、生成 AI は教育現場において学習者の学びや指導者を支援する有効なツールとなる可能性を示すことができた。生成 AI は日々学習し進化を続けており、性能向上によって、今後教育現場においても積極的に活用されることが期待される。

参考文献

- (1) 中央教育審議会(2023) 教師を取り巻く環境整備について緊急的に取り組むべき施策（提言）
https://www.mext.go.jp/content/230828-mext_zaimu-000031605_1.pdf (2024年12月18日最終確認)
- (2) 文部科学省(2023) 初等中等教育段階における生成 AI の利用に関する暫定的なガイドライン Ver1.0
https://www.mext.go.jp/content/20230718-mtx_syoto02-000031167_011.pdf (2024年12月18日最終確認)