

■研究・実践の課題（テーマ）

AI 技術は栄養教育にどのような影響を与えるか

■主任研究者 藤木理代

■共同研究者 福安智哉

■研究・実践の目的、方法、結果、考察や提案等の概要

【序論】

OpenAI 社の ChatGPT が公開されてからその目覚ましいパフォーマンスにより、生成系 AI は凄まじい勢いで社会に浸透しつつある。臨床・研究分野においてもそのパフォーマンスから、利用者の支援ツールとして注目されている。実際、ChatGPT は専門的なトレーニングを行わずとも米国医師国家試験(USMLE)において合格ラインに近い正答率があったことが報告されている (Tifany et al. 2023)。また、日本の医師国家試験(NMLE)においてもチューニングを行うことによって合格点に達することが報告されている。(Tanaka et al. 2023) しかしながら、栄養に関するパフォーマンスに関してはまだまだ不明な点が多い。そこで我々は、ChatGPT と第 37 回管理栄養士国家試験を用いて、ChatGPT の栄養に関するパフォーマンスに関する評価を行う。

【方法】

厚生労働省の公式 Web サイトより第 37 回管理栄養士国家試験問題を入手し、図を利用する問題を削除した計 188 問を問題として利用した。正答においては、厚生労働省の公式 Web サイトより入手し、正答として利用した。その後、3 パターンに分けて判定を行った。

研究 1 : GPT3.5 における栄養に関する知識の評価

先に入手した問題 188 問を ChatGPT3.5 Web 版に入力し、解答を取得し評価を行う。直、プロンプトについては、「回答は数字で行う」ことを求めた。

研究 2 : GPT3.5 プロンプトを使用することによる点数の変化

研究 1 と同様に入力を行うが、プロンプトについて追加を行った。プロンプトについては、「回答は数字で行う」「日本語入力を英語に翻訳し、問題を解答、解答を日本語で出力」「あなたは医学および栄養学の専門家」の 3 点の追加を行った。

研究 3 : GPT 4 における栄養に関する知識の評価

研究 2 において用いたプロンプトを利用し、ChatGPT4 に問題を入力し、研究 2 において取得された結果との比較評価を行う。

【結果】

研究 1 : GPT3.5 における栄養に関する知識の評価

第 37 回管理栄養士国家試験の正答率は 42%であった。科目ごとの正答率は、社会環境と健康 38%、人体の構造と機能 42%、食べ物と健康 24%、基礎栄養学 57%、応用栄養学

73%、栄養教育論 38%、臨床栄養学 19%、公衆栄養学 69%、給食運営管理論 38%、応用力問題 48%であった。

研究2：GPT3.5 プロンプトを使用することによる点数の変化

第37回管理栄養士国家試験の正答率は55%であった。科目ごとの正答率は、社会環境と健康 75%、人体の構造と機能 54%、食べ物と健康 32%、基礎栄養学 64%、応用栄養学 67%、栄養教育論 38%、臨床栄養学 46%、公衆栄養学 54%、給食運営管理論 69%、応用力問題 60%であった。

研究3：GPT4における栄養に関する知識の評価

第37回管理栄養士国家試験の正答率は78%であった。科目ごとの正答率は、社会環境と健康 88%、人体の構造と機能 96%、食べ物と健康 60%、基礎栄養学 100%、応用栄養学 73%、栄養教育論 38%、臨床栄養学 77%、公衆栄養学 92%、給食運営管理論 75%、応用力問題 76%であった。

表：研究1～3における正答率

	研究1	研究2	研究3
	回答率(%)	回答率(%)	回答率(%)
全体正答数	42	55	78
分野別正答率			
社会環境と健康	38	75	88
人体の構造と機能	42	54	96
食べ物と健康	24	32	60
基礎栄養学	57	64	100
応用栄養学	73	67	73
栄養教育論	38	38	38
臨床栄養学	19	46	77
公衆栄養学	69	54	92
給食運営管理論	38	69	75
応用力問題	48	60	76

【考察】

ChatGPT3.5は現在、2022年1月、ChatGPT4に関しては2023年4月までを学習している大規模言語モデルである。そのため、ChatGPT3.5に関しては2023年に行われた第37回管理栄養士国家試験に関する知識を有していない。また、図に関してはまだまだ解釈ができないため今回の研究において図に関する問題は排除した。それらを踏まえて研究1を見ると、特に調整を行わない場合は正答率が5割にも満たないことが分かる。分野別で見ると、臨床栄養学が19%と非常に低い結果を示している。特に症例に関する問題は全問不正解となっており、条件が多くなればなるほど解答が難しくなり、正答率が低下する可能

性を示唆している。しかしながら研究2においては比較的解消されている。これはロールの設定をただけでなく、一度英語化を行うことによって日本語で不明、理解が難しい点が解決され、その結果解答ができるようになったと考えられる。このことは、社会環境と健康、給食運営管理論の正答率が大幅上昇した要因としても考えられる。しかしながら最低限のプロンプト追加を行った研究2においても臨床栄養学は46%と5割を切る正答率であり、臨床分野においてはさらなるプロンプトの追加が必要であると考えられる。

研究3においてはChatGPT4を用いた。研究2と比較してもほぼすべての分野において正答率が増加しただけでなく、国家試験合格点に満たしていた。しかしながら栄養教育論の分野では全く正答率が伸びなかった。プロンプトの追加、およびChatGPT4においても栄養教育論分野が伸びなかった要因としては、生成系AIはあいまいや無難といった「妥当」をまだまだ理解が難しい可能性がある。現段階では栄養教育や臨床分野において生成系AIが利用できるとは言い切ることはできない。しかしながら最低限のプロンプト追加によって正答率が改善されたことから、今後、誤答に関する調査を行い、さらなるカスタマイズによる正答率の増加を行う予定である。