

生成 AI を活用した医療面接練習システムが医学生にもたらす教育効果

小川 歩華 【コミュニケーション & コラボレーション研究室】

Educational Effects of Medical Interview Practice Systems Utilizing Generative AI for Medical Students

OGAWA, Fuka 【Communication and Collaboration Lab.】

1 はじめに

1991年、日本の医学部では臨床実習において参加医学生が十分な知識と技能を備えていることを保証する必要性が指摘され、技能を保証する方法として客観的臨床能力試験 (OSCE: Objective Structured Clinical Examination) が導入された。医療面接は OSCE 試験科目の1つであり、医学的診断に大きく影響する重要な臨床技能である。医療面接の技能習得方法は指導医や他の学生を患者役としたシミュレーション練習がほとんどであった中で、生成 AI の普及によって、生成 AI を仮想患者として練習する方法が提案された。この方法は、学生の医療面接技能の取得に効果があり、従来の方法よりも学生の学習意欲とコミュニケーション力の向上に有効である [1, 2]。生成 AI を用いた医療面接練習は、医療面接教育の主流となることが期待されている。しかし、既存研究の多くは OSCE 前の練習を主な対象としており、応用可能性については十分に検討されていない。よって本研究では、生成 AI を用いた医療面接練習を OSCE 対策にとどめず、多様な教育目的に活用可能な汎用的学習手法へと発展させることを目的とした。

2 症例報告を元にした医療面接シナリオ

2.1 医療面接練習におけるシナリオ作成の課題

医療面接の練習には医療面接用シナリオが必要であるが、シナリオの作成は多忙な指導医にとって時間と手間を要するものである。そこで、シナリオを容易に作成する方法を提案し、指導医の時間的負担の軽減を目指す。

2.2 症例報告を元にしたシナリオ作成方法

生成 AI を活用して症例報告から医療面接用のシナリオを作成する方法を提案する (図 1)。手順 1 では症例報告の情報の整理、手順 2 では情報の創作と追記、手順 3 では情報の具体化を行った。提案方法で作成したシナリオ (提案シナリオ) と従来のシナリオ (従来シナリオ) を用いて生成文章評価を行った。2名の評価者が合計 100件 (提案シナリオ 50件、従来シナリオ 50件) の回答データを1つずつ確認し、各項目 (正確性、関連性、文法、簡潔性、自然さ) 5段階で評価した。また、医学生にアンケートを実施し、提案シナリオの有用性を検討した。

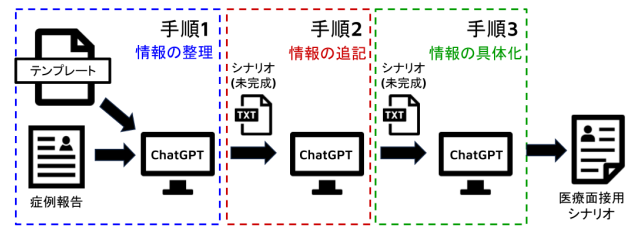


図 1 提案するシナリオ作成方法

2.3 結果・考察

評価の結果、従来シナリオと提案シナリオのどちらも回答は 4.5 以上の高評価となり有意差はなかった。この結果から、提案シナリオが患者パフォーマンスを下げることが無いと言える。アンケートでは提案シナリオも従来シナリオと同程度に継続利用意欲が高く、提案シナリオは医学的知識の深化に、従来シナリオは初心者の自信形成に適していることが示唆された。OSCE で求められる典型症例より症例報告の症例は医学生にとって難度が高いため、研修医や専門医志向の医師を対象にシステムと提案シナリオを組み合わせることで、症例学習教材として有用になると考えられた。また、本手法は 5 分以内でシナリオ作成が可能となるため、指導医の負担軽減と種類豊富なシナリオの大量生成が期待できる。

3 英語医療面接練習の実践

3.1 医学部における英語教育の課題

訪日外国人の増加により国内でも外国人患者を診療する機会が拡大しており、英語による医療面接能力の必要性が高まっている。英語医療面接の学習は語学力向上にとどまらず、患者中心のコミュニケーション能力を養う点でも意義があるが、日本では十分な教育体制が整っていないため、英語医療面接の指導をする教育機関は多くない。そこで、生成 AI を用いた医療面接練習システムを活用した英語医療面接教育を提案する。

3.2 生成 AI を用いた英語医療面接練習

生成 AI を用いた医療面接練習システムで英語練習が効果的に行えるかを検討した。日本語の医療面接シナリオを用いたまま練習言語だけを英語にする。回答をより

実際の患者に近づけるため、カナダ・ウェスタン大学が提供している研究用医療面接データセットを用いて生成AIのモデルを調整した。生成AIの回答が日本人医学生への学習に役立つ自然な英語であることを確認するため、3名のネイティブスピーカーとAIを評価者として2.2と同様の生成文章評価を実施した。

3.3 結果・考察

評価平均は4.5以上と高く、日本語シナリオであっても生成AIは英語話者患者として自然な回答生成が可能であることが示された。モデル調整により簡潔性は向上した一方、自然さについては人間評価で有意差が認められず、評価基準の曖昧さや文化的背景の違いが影響した可能性が考えられた。また、今回調整に用いたデータはOSCEベースの医療面接であり、実際の診療場面の記録ではなく模擬患者相手の会話であったため、回答が標準化されていたことも考えられる。このことから、モデルの調整によって得られる簡潔かつ標準化された回答は、情報を聞き出す能力を鍛える、英語に慣れるという目的を持った学生にとって有用であると考えられた。一方で、既存モデルの回答は想定以上の内容が返ってくるため、相応の英語力が備わっている学生に有用であると考えられた。生成AIの回答を医学英語担当の指導医に見せたところ、「患者回答のレベルが高く学習に有用である」とのコメントを得た。システムでの練習を経験した医学生からは、この練習方法で英語医療表現の理解が深まる、英語医療面接が流暢になるという意見が得られた。このことから、生成AIは英語医療面接技能向上に繋がる練習を提供することができ、現日本医学部の英語医療面接における教育体制の改善が期待できる。

表1 英語生成文章評価

評価者: ネイティブスピーカー			
	既存モデル	調整済モデル	p 値
簡潔性	4.38	4.76	0.000
自然さ	4.85	4.84	0.892
評価者: 生成 AI(gpt-5-mini)			
	既存モデル	調整済モデル	p 値
簡潔性	3.84	4.75	0.000
自然さ	4.51	4.99	0.000

4 医療面接の技能評価の特徴量分析

4.1 医療面接の技能評価における課題

医療面接技能の評価は主観に依存しやすく、構造化された方法、同一の環境で評価を実施しても評価者間にはばらつきが存在する。ゆえに、OSCEにおける評価者間信頼性の確保は依然として課題となっている。したがって、評価者依存性を低減し、多面的に面接技能を定量化できる仕組みが求められる。今回は、会話ログから抽出可能な情報を学年間で比較し、特徴量として有効な指標を見つけ、フィードバックへの応用可能性を検討する。

4.2 会話ログを用いた特徴量分析

医療面接練習システムを用いて収集したデータで特徴量分析を行った。医学部4年生12名から集めた34件のデータと医学部6年生6名から集めた44件のデータから算出した特徴量を比較し、有意差を確認した。収集時期による生成AIの性能差を考慮し、AIの回答内容と回答時間は分析対象から除外した。

4.3 結果・考察

表2は有意差が確認された特徴量である。思考時間は生成AIが回答を送信してから医学生が次の質問を送るまでの時間である。6年生の方が思考時間が短く、面接の手際の良さが伺える。面接1回あたりの思考時間の標準偏差も6年生の方が短く、学生・シナリオ間でばらつきが無く安定していると言える。OSCE型評価とは「挨拶をしている」「主訴を確認している」といったチェックリスト型評価を指している。OSCE評価基準に倣って64個のチェック項目を作成し評価を行ったところ、6年生の方がチェック数が多く、短時間で多く質問できていると言える。これらの指標は医療面接の技能指標の1つとして有用であり、これらの指標から明確な合格基準を設定することで、学習者は基準を意識した効率的な練習が可能になると考えられる。

表2 特徴量の比較結果

特徴量	4年生	6年生	p 値
面接時間	17m40s	10m25s	0.0000
思考時間 (平均)	39s	20s	0.0000
思考時間 (標準偏差)	±34s	±23s	0.0002
OSCE 型評価	30.68	41.16	0.0000

5 まとめ

本研究では、「症例報告を元にした医療面接練習用シナリオの自動生成」と「英語医療面接練習の実践」、そして「技能評価のための特徴量分析」を行い、生成AIを用いた医療面接練習システムの教育ツールとしての可能性拡大を検討した。生成AIが医療面接練習にもたらす教育効果は、医学生の面接技能向上に留まらず、専門医教育、医学英語教育、研修医や医師を対象とした教育にも応用可能であると考えられた。

参考文献

- [1] A. Yamamoto, M. Koda, H. Ogawa, et al. Enhancing medical interview skills through ai-simulated patient interactions: Nonrandomized controlled trial. *JMIR Medical Education*, Vol. 10, No. e58753, p. 12, 2024.
- [2] Z. Wang, TT. Fan, ML. Li, et al. Feasibility study of using gpt for history-taking training in medical education: a randomized clinical trial. *BMC Medical Education*, Vol. 25, No. 1030, p. 11, 2025.