

# 8. 医師国家試験合格水準を持つ 生成AIの開発と今後

岡部 篤史 (株) オルツ新規事業開発責任者 BEO

2023年4月、株式会社オルツとファストドクター株式会社が、オルツが保有する日本発大規模言語処理モデル「LHTM-2」をベースとした人工知能(AI)エンジンをういて、禁忌問題を含む医師国家試験問題の合格基準を満たす生成AIを用いた医療AI(生成系医療AI)を共同開発したことを発表した。ここに、日本初の医師国家試験合格水準のAI「医療AI-LHTM2」が誕生したのである。

2024年4月にも、オルツは最新の医師国家試験に挑戦し、見事合格水準を満たす結果となった。この、初期研修医レベルとも言える生成AIの医療領域における活用可能性について考察する。

## 医師国家試験合格水準の AI活用と可能性

医師国家試験合格水準の生成系医療AIの現状と、海外事例を交えた今後の可能性について述べる。

### 1. テストの考察と現状

2023年2月、医学生が、対話能力が高く無料で使用可能なAI「ChatGPT」(OpenAI社)に医師国家試験の過去問題を解かせたことが話題となった。正答率は55%と合格基準には達しなかったものの、多くの方がAIの情報処理能力を認識した。ChatGPTは特定の分野に特化した学習をしていないにもかかわらず、米国でも医師資格試験や経営学修士(MBA)試験を解かせて合格ラインに達したとの研究が相次いでおり、こうした対話型AIが、将来的には医療現場において医師の診療支援を行うことが期待されている。

2023年度、オルツとファストドクターは度重なる実証実験を行い、医師国家試験の合格基準に到達するAIの開発に成功した。その後、2022~2024年度の医師国家試験の合格基準<sup>1)</sup>をすべて超えた正答率を達成し、禁忌選択問題も含め合格水準を通過する知能を備えていることを示した。

このレベルに到達するに当たっては、インターネット上に出回る、いわゆるデマ情報を排除するため、「医療AI-LHTM2」は医学部の教科書、医学書、

過去問題集、模擬試験の範囲に学習範囲を限定した。また、禁忌問題の正答率向上のため、ファストドクターの臨床医による品質チェックや画像認識を組み合わせることで、禁忌問題の正答率向上を実現した。

しかしながら、課題も残る。年度ごとの総合正答率は82~85%程度となっているが、MRIなどの画像の問題や、質問の短い問題、計算問題などでの間違いが目立った。画像に関しては学習範囲を拡大することによって一定の向上は見込まれるが、一般的な生成AI同様に、質問の短い問題はAIが質問意図を限定できずに間違えたことを述べたり、前提条件が複雑かつ計算中に四捨五入などを求められたりすると、余計なところで四捨五入をし回答数値が若干ずれるなどの課題があった。解答例としては「正: 341mOsm/L」を「誤: 351mOsm/L」や「正: 40mL/分」を「誤: 0.4mL/分」と計算を間違えたりするといった具合だ。

これらの例から、「医療AI-LHTM2」をはじめとした生成系医療AIの間違いが、臨床において致命的なミスを誘引する可能性もあり、現状においては、診断領域での活用は慎重であるべきと考える。

### 2. 米国における生成系医療AI の状況

生成AIの開発は、米国を筆頭に海外勢が先行している。そのため、世界最先端の医療AI活用については、米国における医療AI開発の代表格であるメイヨークリニック(ミネソタ州)の事例を