臨床へのアドバイスサービスの補助的ツールとして Chat GPT を活用する試み

◎村本 美紅[1)]、木村 美香[1)]、川崎 晴希[1)]、寺田 早良[1)]、竹本 賢一[1)]、大江 宏康[1)]
金沢大学附属病院[1)]

【背景】

近年、人工知能（AI：Artificial Intelligence）の進歩により、Chat GPT（Chat Generative Pre-trained Transformer）のような言語処理モデルが注目を集めている。Chat GPT には誤回答、再現性等の問題はあるものの、その柔軟な対話能力から、医療現場における利活用が期待されている。

【目的】

医療現場での Chat GPT 活用の可能性を探るため、血液検査のアドバイスサービスの補助的ツールとして Chat GPT の適用を試みた。

【対象と方法】

2013 年 2 月から 2023 年 3 月において、当院検査部の血液検査室で作成したアドバイスサービス記録 28 件のうち、検査室からの提案 3 件、当院特有の依頼 8 件を除外した 17 件（血算 13 件、凝固 4 件）のアドバイスサービス事例を対象とした。

Chat GPT は、Open AI が提供している Chat GPT-3.5 のフリープランを使用した。

17 件のアドバイスサービスについて、以下の独自の方法を用いて評価した。

①臨床側からの依頼内容を Chat GPT に入力し、モデルからの回答を収集した。

②収集した回答を実際のアドバイスサービス提供者 2 名（いずれも血液検査の経験年数 16 年）が 3 段階のスコア（表 1）によって評価し、2 名の合計スコアから適切性や有用性を判断した。③さらに、合計スコア 4 点が得られなかった回答 13 件については、依頼内容を具体的な内容に変更して入力を行い、収集した回答を再評価した。

【結果と考察】

結果を表 2 に示す。合計スコア 0、2、4 点については、アドバイスサービス者 2 名のスコアが一致していた（一致率 50%）。合計スコア 1、3 点は両者のスコアが不一致であったが、スコアは全て 1 点差であった。また、再収集を行った 13 件のうち、11 件で再評価時のスコアが上昇していた。再評価を行った 13 件の初回平均スコアが 1.85 点であったのに対し、再評価の平均スコアは 3.31 点に上昇した。

入力内容が具体的なものほど、高スコアになる傾向が見られた。低スコアのものは入力内容を工夫することで改善された。ただし、一部の回答には内容に矛盾がみられたため、代替手段として用いるには適切とは言えなかった。

【結語】

血液検査のアドバイスサービス事例に Chat GPT の適用を試みた。一部の事例で誤回答が見られたため、アドバイスサービスにおける活用には慎重な検討が必要である。

表 1

スコア	評価
2	説明として十分である
1	説明として一部不十分である
0	説明として不十分である

表 2

合計スコア	初回評価（n=17）	再評価（n=13）	評価
4	4件	6件	説明として十分である
3	5件	5件	説明として一部不十分である
2	3件	2件	
1	3件	0件	
0	2件	0件	説明として不十分である

連絡先：村本美紅（076-265-2000、内線 7180）