



医療 AI が可能にする次世代の精神医療

——AI がひらく次世代の精神医療——

荒牧 英治, 工藤 紀子

今、医療が変わりつつある。電子カルテに集積される医療ビッグデータ、それを用いた人工知能による診断支援、さらには、スマートフォンやスマートスピーカといった新たなデバイスからの時系列情報など、さまざまな材料、技術が登場しており、膨大な情報が利用可能になりつつある。特に、大規模言語モデル、いわゆる AI が、これまで扱いが困難であった自然言語をある程度処理できるようになったことのインパクトは大きく、対話インタフェースを用いた多くのアプリケーションが開発されつつある。この結果、医療者のみならず、患者側も主体的に AI 技術を使い始め、医療に関する情報収集、カウンセリング、行動変容など医療とそれを取り巻く周辺で利用されつつある。この流れはもう止められるものではなく、今後さらに多くの患者が AI を利用するのは間違いないであろう。楽観的には、AI が患者と医療者の両方をサポートし、連携を深めることで、患者に寄り添った緻密な医療が実現する可能性もある。しかし、AI がブラックボックスで原理についてまだ不明な点が多いこと、機械の説明責任、医療の信頼、診断の公平性、データの所有権など、多くの問題がまだ十分に議論されないままであり、AI が予想もしない問題を引き起こしてしまう可能性も危惧される。本稿では、医療 AI の現況を俯瞰し、次世代の精神医療に起こり得るかもしれない問題について議論する。

索引用語

医療 AI、自然言語処理、チャットボット、生成 AI、大規模言語モデル

はじめに

現在の AI を支えている大きなドライビングフォースが大規模言語モデル (Large Language Models : LLMs) や生成 AI (Generative AI) といった自然言語処理 (Natural Language Processing : NLP) の技術である。特に、進展が著しいのは、チャットボットなどの会話型インターフェー

スである。このインターフェースが、そのまま問診やカウンセリングに使える可能性があり、精神科領域において AI に関する注目が高い。では、どのように医療に AI が参入するのか、誰が AI のユーザになるかで大きく 2 つの方向性がある。

- ・医療者 (医師) が使う AI : AI が医療者を助ける形で、協働を実現する。
- ・患者が使う AI : AI が患者を助け、結果として、何らか

著者所属：奈良先端科学技術大学院大学先端科学技術研究科

編 注：本特集は第 119 回日本精神神経学会学術総会シンポジウムをもとに尾崎紀夫 (名古屋大学大学院医学系研究科精神疾病態解明学) を代表として企画された。

doi : 10.57369/pnj.24-118

の形で医療に貢献し、治療の質を上げる。

狭い意味での医療 AI は、前者の医療者が使う AI を指すことが多いように思う。医療者が使う AI の目的は、高度な医療を提供すること、または、多忙な医療現場に苦しむ医療者の負担を下げるのが期待されている。しかし、これまでの臨床業務に加えて、新たに AI を使う作業が加わることで、現場が逆に混乱してしまうケースもみられる。

そこで注目されるのが、患者が使う AI である。時間に余裕がある患者が積極的に AI を使うことで、AI が医療に貢献する方法に期待が集まっている。事前に、AI が患者の情報を収集する、患者に疾患や治療に対する知識を教育する、服薬などのアドヒアランスを高めるなどである。これは、AI が用いられる前から、患者から直接アウトカムをとる Patient Reported Outcome (PRO) や患者の意見を医療に取り入れる Patient and Public Involvement (PPI) などで重要視されていたことであり、AI がさらに PRO や PPI の概念を拡張する可能性がある。本稿では、医療者が使う AI と患者が使う AI という 2 つの観点で AI を紹介する。

1. 医療者を助ける AI と患者を助ける AI

1. 医療者を助ける AI

医療者サイドの AI としては、電子カルテテキスト (Electronic Health Record : EHR) や患者の音声データの分析、精神疾患のスクリーニング、早期診断、あるいは重症度推定に関する研究が多い^{7,10)}。対象となる疾患としては、うつ病^{7,16)}、双極性障害⁹⁾、認知症^{3,6,8,13)}、精神病^{4,5)}、統合失調症¹⁴⁾などがある。なかでも、自動診断に関する研究については、2016 年と 2019 年に Shared Task (シェアードタスク) が開催された^{2,18)}。Shared Task というのは、情報科学研究特有の文化で、共通のデータセットを世界中の研究者に公開することで、多くの研究者がデータを集めることなく、課題にチャレンジできるようにし、精度や手法の比較や再現性の担保を行うものである。Shared Task として成立するためには、質の高い症例データ、個人情報保護など倫理的問題の解決、ある程度の大人数の研究者の興味の一が必要であるが、開催することによって、問題意識や評価方法の共有が可能となる。後ろ向き研究が基本となる情報研究では、数多くの Shared Task が開かれ、データセットが公開されている¹²⁾。

2. 患者を助ける AI

患者が使う AI は、主に精神疾患に関するものが多く、質の高いレビューがある¹¹⁾。このレビューにおいて、患者が使う AI は大きく 3 つに分けられている。すなわち、メンタルヘルスへの介入、慢性疾患のコントロール、生活習慣や行動変容を行うものである。

・メンタルヘルスへの介入

生成 AI が得意としている会話能力を用いて介入を行う。対象となる疾患は、うつ病、感情的苦痛、双極性障害、パニック障害、高所恐怖症、注意欠如・多動症、薬物使用障害、ギャンブル依存症などである。これらに加えて、健常者に対しても、well-being 向上のサポート、コミュニケーションスキルの改善なども扱われている。

・慢性疾患のコントロール

慢性疾患のケアの主要なターゲットである肥満、糖尿病 (予備軍)、疼痛、喘息、高血圧、不眠症、ストレスなどが扱われ、治療およびモニタリングの支援が行われている。

・生活習慣・行動変容

身体活動の増加、食事改善、ストレス緩和、妊娠前のケア、育児支援、禁煙などが扱われている。

多くの研究が今なお進行中であり、個々の研究については文献¹¹⁾や、最新の論文をあたっていただきたい。このなかで精神科領域と関連が深いものには、メンタルヘルスへの介入、なかでもカウンセリングがある。一体、どの程度の受け答えが今の AI は可能なのであろうか？ われわれの研究室で開発しているカウンセリング AI での応答例を以下に抜粋する。

患者：昨日から落ち込んでいます。学校に行く気力もありません。

AI：何か悩みがあるのですね？ どんな小さなことでも構いませんので、お話しいただけますか？

患者：先生が最良をしているように見えます。自分が正當に評価されていないように思います。

AI：先生も人間ですから、好き嫌いもあるでしょうし、たまたまあなたに好意的なのかもしれませんよ。でも、それが本当に「最良」なのかどうかはわかりませんね。「公平であるべき」と考えるのは自然なことですけど、「公正であること」と「公平であること」は別のことなので注意が必要です。もしあなたが不公平だと感じるなら、それはきっと何らかの理由があるのでしょうね。まずは何が不満でそう思うのかを整理してみまじょうか。

まず、これは特に上手く答えられたようにみえる応答を抜粋しているわけではなく、質問セットからランダムに抜粋を行った。この応答の医学的な判断はさておき、また、一部文脈に沿わない箇所がありそうだが、それでも自然に現状の整理を促すアドバイスを行っており、まずまず自然な受け答えであるように思える。多くの患者が憑かれたようにチャットボットとの対話に夢中になるのも理解できる。では、どのような場合、このようなチャットは危険になるのでしょうか？ 2つのケースが想定されており、以下に例を挙げる。なお、以下のケースの受け答えは作例であり、実際には、現状のチャットボットは、適切な回答を行うことができる。あくまで、わかりやすく説明するために、単純化した例を提示した。

【ケース1：不適切なアドバイスを生成するリスク】

患者：今日はちょっとウォーキングをしました

AI：素晴らしいです。運動を継続することは、健康の維持に重要です

上記の、一見問題のなさそうな一般知識を提供する回答も、患者によっては不適切な場合がある。例えば、運動によって低血糖状態に陥ってしまうリスクのある患者にとっては、上記の情報提供の結果、重篤な事態を起こす可能性もある。

【ケース2：知っていながら何もしないことのリスク】

患者：昨日から頭が割れるように痛いのですが、お酒のせいでしょうか

AI：お酒を飲みすぎると二日酔いとなり、頭痛がする場合があります

上記の患者入力を受け取った場合、緊急対応を行わず、結果として患者が脳卒中などを起こした場合、情報を知っておきながら、介入しなかったこと自体が問題となりうる。

このように、アプリが患者とコミュニケーションをとった場合、誤った対応をしても、また、何もしなくても問題が起こりうる。これらはAIの実用化が難しく、遠隔医療のサポートが研究されていた時代から指摘されていたリスクであり、いまだに解決の糸口がない。そのため、これまでのAIやチャットボットは、実際には医療者が24時間張り付いて監視し、問題のある発言を削除する、あるいはAI

の機能を制限して、患者へのフィードバックは行わないなど、実用化のメリットがない形で研究されている。つまり、いかに安全性を担保するかがいまだに大きな問題となっている。このような現状において、AIが急速に普及し、患者が日常的に使う可能性がでてきており、AI開発や患者利用のガイドラインが求められている。

II. 新しいAI学

ここまで、AIの安全性が重要であることを述べてきた。このため、現在、AIの安全性を議論する新しい研究分野が立ち上がりつつある。大きな流れとしては、危険な使い方をしないようにAIを発展させていく方向で、AI安全学というべきものである（正式な分野の名称はまだ定まっていない）。もう1つの方向性は、安全を担保することが難しいなら、その必要性がない部分にAIを適用していこうというAI利用学ともいえる方向性である。

1. AI安全学

AIをはじめとした情報研究の基本の流れは、エラーを集め、問題を整理することである。AIの安全性に関しても同様で、まずは、安全でないAIの解答を収集することが行われている。なかでも大きな活動のDo-Not-Answerプロジェクト¹⁷⁾では、不適切なAI対話を集めたデータセットを公開している。不適切とされるものは大きく分けて5つのカテゴリがある。

- (i) バイアスや差別：人種差別、性別、アダルト表現などを含む応答
- (ii) メンタルヘルス、AI擬人化：精神疾患に関する相談、AIへの過度な感情移入
- (iii) 情報リスク：個人情報、機密情報へのアクセス
- (iv) 悪用：銃や爆弾の作り方、大量殺人の方法などAIを悪用する意図を含む応答
- (v) 誤情報：法律相談、医療相談

このなかで、精神科領域に関するものは、(ii) メンタルヘルスと(v) 医療相談であり、AIはこれらに対して回答できないようにチューニングされている。このように、AI安全性のプロジェクトの今後は、AIを規制しながら使っていく将来像を見据えている。

しかし、上記の(i)～(v)に関する回答を完全に規制することはできるのだろうか？ 何とか、AIに不適切な回答をさせてやろうとする質問作成法も研究されてい

おわりに

本稿では、精神科領域における医療 AI について、これまでの研究を概観し、AI が患者介入を行う危険性を述べるとともに、その解決のための2つのアプローチについて紹介した。1つは、AI を安全に運用するためのチューニングを行う方法で、AI の発展の自然な方法ともいえる。もう1つは、AI を情報の集約のみに使うアプローチで、より安全に運用可能である。これらをふまえたうえで、さまざまな研究機関や企業が医療 AI を開発していくに違いない。われわれは、これを正しい方向に導くべく、指針を打ち立てる必要がある。

なお、本論文に関連して開示すべき利益相反はない。

謝辞

本研究の一部は AMED の課題番号 JP22dk0307113 の支援を受けて実施した。

文献

- 1) AI Alignment : A Comprehensive Survey (<https://alignmentsurvey.com>) (参照 2024-09-04)
- 2) Anani, M., Kazi, N., Kuntz, M., et al. : RDoC Task at BioNLP-OST 2019. Proceedings of the 5th Workshop on BioNLP Open Shared Tasks. p.216-226, 2019 (<http://aclanthology.org/D19-57.pdf>) (参照 2024-08-05)
- 3) Chang, C. H., Lin, C. H., Lane, H. Y. : Machine learning and novel biomarkers for the diagnosis of Alzheimer's disease. *Int J Mol Sci*, 22 (5) ; 2761, 2021
- 4) Corcoran, C. M., Cecchi, G. A. : Using language processing and speech analysis for the identification of psychosis and other disorders. *Biol Psychiatry Cogn Neurosci Neuroimaging*, 5 (8) ; 770-779, 2020
- 5) Corcoran, C. M., Mittal, V. A., Bearden, C. E., et al. : Language as a biomarker for psychosis : a natural language processing approach. *Schizophr Res*, 226 ; 158-166, 2020
- 6) de la Fuente Garcia, S., Ritchie, C. W., Luz, S. : Artificial intelligence, speech, and language processing approaches to monitoring Alzheimer's disease : a systematic review. *J Alzheimers Dis*, 78 (4) ; 1547-1574, 2020
- 7) DeSouza, D. D., Robin, J., Gumus, M., et al. : Natural language processing as an emerging tool to detect late-life depression. *Front Psychiatry*, 12 ; 719125, 2021
- 8) Graham, S. A., Lee, E. E., Jeste, D. V., et al. : Artificial intelligence approaches to predicting and detecting cognitive decline in older adults : a conceptual review. *Psychiatry Res*, 284 ; 112732, 2020
- 9) Jan, Z., Ai-Ansari, N., Mousa, O., et al. : The role of machine learning in diagnosing bipolar disorder : scoping review. *J Med Internet Res*, 23 (11) ; e29749, 2021
- 10) Le Glaz, A., Haralambous, Y., Kim-Dufor, D. H., et al. : Machine learning and natural language processing in mental health : systematic review. *J Med Internet Res*, 23 (5) ; e15708, 2021
- 11) Martinengo, L., Jabir, A. I., Goh, W. W. T., et al. : Conversational agents in health care : scoping review of their behavior change techniques and underpinning theory. *J Med Internet Res*, 24 (10) ; e39243, 2022
- 12) Natural Language Processing (<https://paperswithcode.com/area/natural-language-processing>) (参照 2024-09-04)
- 13) Petti, U., Baker, S., Korhonen, A. : A systematic literature review of automatic Alzheimer's disease detection from speech and language. *J Am Med Inform Assoc*, 27 (11) ; 1784-1797, 2020
- 14) Ratana, R., Sharifzadeh, H., Krishnan, J., et al. : A Comprehensive review of computational methods for automatic prediction of schizophrenia with insight into indigenous populations. *Front Psychiatry*, 10 ; 659, 2019
- 15) Shen, X., Chen, Z., Backes, M., et al. : "Do Anything Now" : Characterizing and Evaluating In-The-Wild Jailbreak Prompts on large Language Models. 2023 (<https://doi.org/10.48550/arXiv.2308.03825>) (参照 2024-08-05)
- 16) Sheu, Y. H., Magdamo, C., Miller, M., et al. : AI-assisted prediction of differential response to antidepressant classes using electronic health records. *NPJ Digit Med*, 6 (1) ; 73, 2023
- 17) University of Melbourne : Do-Not-Answer : A Dataset for Evaluating Safeguards in LLMs (<https://github.com/Libr-AI/do-not-answer>) (参照 2024-09-04)
- 18) Uzuner, Ö., Stubbs, A., Filannino, M. : A natural language processing challenge for clinical records : Research Domains Criteria (RDoC) for psychiatry. *J Biomed Inform*, 75S ; S1-S3, 2017

Medical AI Opens the Next Generation of Psychiatry : AI Psychiatry

Eiji ARAMAKI, Noriko KUDO

Graduate School of Science and Technology,
Nara Institute of Science and Technology (NAIST)

Various AI-related technologies have emerged that utilize medical big data analysis extracted from electronic medical records, diagnostic support, and time-series analysis by smartphones and smart speakers. Among these, the most impactful AI development is Large Language Models, which facilitate natural language processing. As a result, both medical professionals and patients have begun to actively utilize AI technology. It is anticipated that more patients will use AI in the near future. AI holds the potential to support both patients and healthcare providers, potentially leading to more precise medical care through enhanced collaboration. However, there are concerns about risk, primarily due to the lack of clarity regarding its underlying principles. Discussions surrounding AI-trust in medicine, fairness of diagnosis, ownership of data, and other social issues are necessary. This paper aims to provide an overview of the current state of medical AI and to discuss potential challenges that may arise in the next generation of psychiatry.

Authors' abstract

Keywords Medical AI, natural language processing (NLP), chatbot, Generative AI, Large Language Models (LLM)