

厚生労働科学研究費補助金（がん対策推進総合研究事業）
（分担研究報告書）

LLMファインチューニングを用いた相談対応支援ツールの開発

研究分担者 内村祐之 東京科学大学病院医療情報部特任講師

研究要旨

先行研究班（厚労科研20EA1015）で開発したテキストマイニングを用いた相談支援内容の解析ツールを用い診療科ごとに生じやすい相談/対応内容の可視化と課題の分析を行っていたが、昨今の急速な生成AIの進化があることから、班内で検討の結果、現段階で最も有用と考えられる相談対応内容の可視化の検討に用いるIT技術での検討を行うこととした。相談支援センターの活動補助・支援につなげるツールの開発を行うことを目的とし、ChatGPT等の大規模言語モデル（LLM）を用いた相談対応支援ツールの開発に向けた検討を行うこととした。また、相談内容を一切外部に出さずに利用可能なローカルLLMによるツールの検討も併せて行う事とした。

A. 研究目的

現在、全国のがん相談支援センターでは、一定の基準を満たした上で治療の内容や設備、がんに関する情報提供などを行っており、がん相談支援における相談内容をデータとして管理している。しかし、蓄積されたデータの利活用や相談員間の知識共有は十分に出来ていないのが現状である。将来の相談内容の更なる多様化に備えるためには、蓄積された相談情報を活用し、FAQの作成や相談員間の情報交換など他施設を含めた相談員のノウハウを共有する仕組みを作り、相談員による経験や知識に起因する対応の差を埋めることにより、一定水準以上の相談対応が行えるような相談対応支援の仕組みの構築が必要であると考えられる。先行研究班（厚労科研20EA1015）で開発したテキストマイニングを用いた相談支援内容の解析ツールを用いて、診療科ごとに生じやすい相談/対応内容の可視化と課題の分析を行ってきたが、昨今の急速な生成AIの進化があることから、班内で検討の結果、現段階で最も有用と考えられる相談対応内容の可視化の検討に用いるIT技術での検討を行うこととした。まずは相談支援センターの活動補助・支援につなげるツールの開発を行うことを目的とし、ChatGPT等の大規模言語モデル（LLM）に実際のがん相談に関する追加学習を行ったモデルを用いた相談対応支援ツールの開発に向けた検討を行うこととした。またChatGPTの利用は、相談対応データが外に出たまま医療機関が安心して利用できる環境ではないことから、院内での生成AI活用環境の検討も併せて行う事とした。

B. 研究方法

初めに、OpenAI社が2024年7月18日に発表した生成AI「GPT」シリーズの新モデル「GPT-4o mini」を使った検証を行った。GPT-4o miniは「ChatGPT」を提供するOpenAIが発表した新しい生成AIモデルである。GPT-4o miniはコストパフォーマンスを重視

した軽量な生成AIモデルで、これまでのGPT-3.5 Turboよりも回答精度も高く、知識のカバー期間も2023年10月までとなっているのが特徴である。拠点病院2施設の相談記録100件（個人情報を除いたデータ）を研究資料として活用し、教師データとしてファインチューニングに利用した。相談内容、対応内容共に複数の内容が含まれているものも1つの内容として学習データを作成し、【相談内容】、【対応内容】のラベルも全てのデータに共有して付与を行った。以下に学習データの一部を示す。

```
{ "messages":  
  [  
    {  
      "role": "system",  
      "content": "このチャットボットはがん相談に関する質問に答えます。"  
    }, {  
      "role": "user",  
      "content": "【相談内容】 本人来室。医療費が高いと思う。内訳を確認してか  
    }, {  
      "role": "assistant",  
      "content": "【対応内容】 高額療養費制度について説明。ひと月は暦月なの  
    }  
  ]  
}
```

次に、ChatGPTの利用は相談対応データが外に出たまま医療機関が安心して利用できる環境ではないことから、院内での生成AI活用環境の検討（ChatGPTを用いるのと同様の環境を院内パソコン1台でできるか）の検証を行った。

（倫理面への配慮）

データは全て実データを元にしたサンプルデータを用いた。実証試験では実証試験施設である青森県立中央病院の倫理審査委員会の承認を経て、倫理規定等を遵守しながら進めている。

C. 研究結果

ChatGPTのファインチューニング（Fine-tuning：微調整）とは、あるデータセットを使って事前学習（Pre-training）した訓練済みモデルの一部もしくは全体を、別のデータセットを使って再トレーニングすることであるが、今回はGPT-4o miniのモデルに対して相談記録100件を使ったファインチュー

ニングを行った。ファインチューニング前のGPT-4 o miniに「地域住民より電話相談あり。胃がん検診を受けたが、要検査の知らせが届いた。医療機関の一覧が入っているが、どこにかかればよいのか？がんだったらどうしよう、最悪の結果だったらと思うと不安で眠れないがどうしたらいいか？」と質問を

ASSISTANT

胃がん検診の結果が要検査である場合、まずは医療機関の一覧を見て、信頼できる胃がんの専門医がいる病院やクリニックを選んで受診することを「お勧めします。胃がんは早期発見が重要ですので、専門家の指導のもとで迅速に対処することが大切です。

不安や心配があるのは当然のことですが、その不安をしっかりと相談することで少しでも軽減できることもあるかもしれません。また、検査結果に従って医師の指示に従い行動することが大切です。

して得た回答の一部を以下に示す。回答に曖昧な部分が多かったが概ね質問に沿った回答が返ってくる事は確認できた。次にファインチューニングを行った後に、同様の質問を投げて得られた回答を以下に示す。

(回答)「不安なお気持ちは当然であることをお伝えし、一旦お話を傾聴。胃がんの検査から診断までの流れについて情報提供。まずは必要な検査を行う必要があるため、医療機関一覧から受診できそうな所を選定し、受診していただくようお話。今後も引き続き相談可能であることをお伝えした。」

次にChat-GPTと同様の事がインターネットに接続していない院内のPC1台を用いローカルLLMの環境が構築出来るのかの検証を行った。構築した環境は以下の通りである。

PC	Apple MacBookPro
CPU	Apple M2 Max
メモリ	32GB
LLMモデル	ELYZA-japanese-Llama-2

ローカルLLMでの検証結果

(質問)「東京で乳がんのセカンドオピニオンを受けたい。どうすればいい？」

【回答】

1. 乳がんのセカンドオピニオンを受けたいと考えていることを伝える - 電話やメールで問い合わせる際に、事前に相談したい内容や悩みを伝えると良いです。
2. 希望する日時や診療科を伝える - 予め自身が希望する診療科や、受けたい検査や治療法などを伝えることで、スムーズな相談ができます。
3. 乳がんのセカンドオピニオンを受けられる医療機関を調べる - 東京には多くの乳がんのセカンドオピニオンを受けられる医療機関があります。自身に適した医療機関を調べ、予約を取りましょう。
4. 乳がんのセカンドオピニオンを受ける - 乳がんのセカンドオピニオンを受ける際は、問診表や健康診断の結果書、レントゲン写真などを持参しましょう。
5. 相談内容の確認と説明 - 相談内容について説明を受け、納得がいくまで質問や確認をしましょう。
6. 治療方針の決定 - 相談内容によっては、その場で治療方針の決定をする場合があります。
7. 乳がんのセカンドオピニオンを受けた後のフォロー - 乳がんのセカンドオピニオンを受けた後は、定期的な検診やフォローアップを受けることが大切です。

乳がんのセカンドオピニオンは、自身の病状や治療方針について確認をするためのものです。治療方針を決定する際には、様々な視点から判断をする必要があります。そのため、専門家の意見を聞くことは患者様にとって重要です。乳がんのセカンドオピニオンを受けることで、より適切な治療方針を決定することができるでしょう。

D. 考察

このように昨今の生成AIにおいては、一般的な質問に関しては概ね質問に沿った情報が返ってくるが、実際に現場で相談員が日々対応しているがん相談は専門性が高く、生成AIがインターネットに公開されている情報の学習だけではした適切な内容を得るのは精度に課題あることが予見できたが、実際に検証を行った相談に対する回答(対応内容)の検討では、相談対応しているような回答は得られるものの精度にはやはり課題が多くあることが示された。ファインチューニングそのもののデータの取り込みや方法についてはまだ改善の余地が大きいと考えられたことから、引き続き活用する情報源などを精査しながら改善を図り検討を継続する予定である。またChatGPTの利用は、相談対応データが外に出てしまい医療機関が安心して利用できる環境ではないことから、インターネット接続の必要のないローカル環境のPCで動作するLLMによるツールの検討も今後行っていく予定である。

E. 結論

ChatGPTのファインチューニングを用いた検証と併せて、院内での生成AI活用環境の検討

(ChatGPTを用いるのと同様の環境を院内パソコン1台でできるか)を行ったが、その結果ほぼ同様の環境がつくることが確認できた。今後は、「がん情報サービス」や「がん診療ガイドライン」等の信頼できる情報をファインチューニングに用いた簡易版生成AI活用ツール開発・作成し、院内環境で利用できるか等の検討を行う予定である。

F. 健康危険情報

特になし

G. 研究発表

(発表誌名巻号・頁・発行年等も記入)

1. 論文発表

なし

2. 学会発表

なし

H. 知的財産権の出願・登録状況

(予定を含む)

1. 特許取得

なし

2. 実用新案登録
なし

3. その他
なし

