

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）

がん相談支援の質の確保及び持続可能な体制の構築に資する研究

テキストマイニングツールとLLM（大規模言語モデル）を用いた相談支援内容の解析と
相談支援センターのフィードバックを得る方法の検討

研究分担者 内村祐之 東京医科歯科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に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い、更にLLM（大規模言語モデル）の生成AI機能をツールに組み込み、がん相談支援業務の対応における相談員の業務負担軽減の可能性を検討する。また、相談支援センターのフィードバックを得る方法について課題を洗い出した上で実現可能な方策を検討する。

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いるが、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。仕組みの構築に向けて、がん相談支援で過去に対応した相談記録情報を利用し、テキストマイニング技術により疾患別やカテゴリー別の傾向を分析し可視化する為、さらに、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う。更に、LLM（大規模言語モデル）の生成AI機能をツールに組み込み、がん相談支援業務の対応における相談員の業務負担軽減の可能性を検討する。また、相談支援センターのフィードバックを得る方法について課題を洗い出した上で実現可能な方策を検討する。

B. 研究方法

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、LLM（大規模言語モデル）の生成AI機能をツールに組み込む。LLMについて今年度はChatGPT（gpt-3.5-turbo-0125）を用い、がん相談に関するサンプルデータ10件にてファインチューニングを行い、既に学習済みのモデルに新たな層を追加し、モデル全体を再学習させた上で使用した。また、相談支援センターのフィードバックを得る方法について会議への参加や病院へ訪問し、フィードバックを実現するための課題についてヒアリングを行なった。

C. 研究結果

1. LLMのツールへの組み込み（10件のサンプルデータを用いたChatGPT（gpt-3.5-turbo-0125）のファインチューニング結果）

① サンプルデータ（抜粋）

相談内容	対応内容
地域住民より電話相談あり。乳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とのご相談あり。	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。乳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願い。今後も引き続き相談可能であることをお伝えした。

② ファインチューニング用データ（抜粋）

```
{ "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": "地域住民より電話相談あり。乳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とのご相談あり。" }, { "role": "assistant", "content": "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。乳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願い。今後も引き続き相談可能であることをお伝えした。" } ] }
```

③ ファインチューニング前後の生成AI結果

相談内容	生成AI結果	
	ファインチューニング前	ファインチューニング後
地域住民より電話相談あり。乳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のご相談あり。	まずは、お気持ちをお察しいたします。乳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。(以下略)	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。乳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願い。今後も引き続き相談可能であることをお伝えした。
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？	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。(以下略)	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願い。今後も引き続き相談可能であることをお伝えした。

2. 相談支援センターのフィードバック検討の為の情報収集

東京都がん診療連携協議会第37回相談・情報部会担当者連絡会にオブザーバとして参加、また、青森県立中央病院、山形県立中央病院を訪問し、相談支援センターのフィードバック検討の為の情報収集を行なった。フィードバックについては患者満足度調査に組み込んで行なっている施設やQRコードを使って行っている施設もあったが、院内調整の難しさや限られた人員で行なっている相談業務への負荷などの課題が明らかとなった。

D. 考察

10件の相談記録情報サンプルデータを用いたLLMの生成AIの結果からはファインチューニングの前後で生成される文章が明らかに異なり、追加学習の効果が確認できた。学習データで「乳がん」となっていた相談内容を「胃がん」に置き換え、表現を変えて質問したが、生成AIから生成される結果は学習データと比較して正しいものであった。ファインチューニング前は、がん相談に関する学習データが限定的だったために、当たり障りのない一般的な内容での生成に止まっていたが、サンプルデータを追加学習させることで、がん相談における相談内容に対応する対応内容が生成される事が確認できた。今回、ChatGPTのファインチューニングで追加学習を行なったサンプルデータは10件分のがん相談の相談内容と対応内容であったが、元データを詳しく見ると、一件分のデータに幾つかの相談内容が含まれる事が確認できたので、今後ファインチューニングを行う際の追加学習用データ作成の際にはデータの分割も検討する必要があると考えた。また、サンプルデータの件数も10件のみであったため、ファインチューニングで学習させる件数について更に増やすことも検討したい。

E. 結論

今年度はChatGPTを使ったLLMでの検証であったが、次年度以降はローカル端末で動作するLLMの検証を加え、がん相談支援にLLMを活用する際のLLMの選定やファインチューニングについて更なる検証を行い、がん相談支援業務の相談員の業務負担軽減の可能性を検討していきたい。相談支援センターのフィードバック検討の為の情報収集においては、現状の把握と幾つかの課題を明らかにする事ができたので、次年度以降具体的な検討を進めていきたい。

F. 健康危険情報

特になし

G. 研究発表

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