

2023-A-22 がん情報提供のためのチャットボット（自動応答）システムの構築に向けた検討

八巻知香子

がん対策研究所 がん情報提供部

研究の分野名

社会医学・国際戦略分野

研究の概要

第3期がん対策推進基本計画では、医療技術や情報端末が進歩し、がん患者の療養生活が多様化する中で、患者やその家族の治療上の疑問や、精神的・心理社会的な悩みも多様化していることから、拠点病院等や小児がん拠点病院のがん相談支援センターが中心となって相談支援の一層の充実をすべきことがあげられた。同計画の中間評価においては、「あらゆる分野で、情報提供及び普及啓発の更なる推進が必要であること」とされ、令和4年に改訂された、がん診療連携拠点病院等の整備にかかる指針においては、「更なる相談支援体制の整備を推進するために、「外来初診時から治療開始までを目処に、がん患者及びその家族が必ず一度はがん相談支援センターを訪問することができる体制を整備することが望ましい」ことを盛り込まれている。現在検討が進められている第4期がん対策推進基本計画（案）でも、より一層の相談支援体制の充実の必要性が記載されるとともに、ICTの活用、真偽の定かでないインターネット情報についての注意喚起の必要性も指摘されている。すなわち、患者や市民が必要とするがん情報の範囲は広がり、情報量は増え、流布する情報も飛躍的に増える中で、その人自身が必要とする適切な情報にたどり着くことがより重要となる一方、困難さは増し、相談対応のニーズも増大している。がん診療連携拠点病院のがん相談支援センターは誰にでも情報を提供し、相談支援を行う窓口であり、「一度は訪問できる」状況となることが望まれているが、対応できる人的資源に限られる以上、ICTを活用したシステムにより自分で適切に情報を得ることが容易にできる仕組みもまた必要である。

本研究に先立って行われた「がん情報提供の自動化に向けた研究（2020-A-19）」では、「大量のデータより患者・家族の情報ニーズを収集し、解析する手法を検討すること」「チャットによるがん相談システムの導入手法の検討」を行った。その結果、（1）人が対応する文字による相談である「チャット相談（有人チャット）」は、電話相談、対面相談とは異なるニーズに対応できること、（2）手動で作成したFAQの積み上げを基盤としたシナリオ型チャットボットに、一部のAI型チャットボットを組み合わせることで、情報提供型の相談については対応可能な内容が多数あると想定されること、（3）有人チャットのデータ（文字のやりとり）はチャットボットの教師データとして直接活用可能であることが明らかとなった。

そこで、本研究では、信頼を確保しつつ効率的にがんの情報をより多くの人に届けるために、新しい情報技術を用いた情報提供の自動化について検討し、実用可能なチャットボットの仕様確定を行うことを目的とする。不安や精神的な苦痛等は対人での対応が必須であるが、情報を求める相談に対してチャットボットでの情報提供ができることになれば、待ち時間なくいつでも活用でき、夜間や休日などがん相談支援センターが稼働していない時間には特に有用と考えられるほか、性的なこと等、対人では相談しづらい内容についての情報提供には特に有用であると考えられる。

また、人が対応するチャット相談と、チャットボットによる相談（情報提供）のデータは、システムの学習データとして蓄積されていくことから、活用される中でシステムが自己学習する仕組みを取り入れることで、ニーズを継続的に把握し、より適切な情報提供を行うインターフェイスとして充実していくことが期待できる。

研究経費

7,312千円

研究班の組織

研究者名	所属研究機関名 所属部局 職名	分担研究課題名
八巻知香子	国立がん研究センター がん対策 研究所がん情報提供部 患者市民 連携推進室 室長	全体統括
堀抜 文香	国立がん研究センター がん対策 研究所がん情報提供部たばこ政策 情報室	相談データの持続的な蓄積方法の検討
内村 祐之	東京医科歯科大学 東京医科歯科 大学病院 医療情報部 特任講師	提示情報の持続的な更新方法の検討
安齋 達彦	東京医科歯科大学 M&Dデータ科学 センター 講師	提示情報の持続的な更新方法の検討
荒牧 英治	奈良先端科学技術大学院大学 先 端科学技術研究科 教授	チャットボットシステムの構築
余 宛柔	奈良先端科学技術大学院大学 先 端科学技術研究科 特任助教	チャットボットシステムの構築
藤牧 貴子	奈良先端科学技術大学院大学 先 端科学技術研究科 研究員	チャットボットシステムの構築

研究の目的と到達目標及び実績要点

全期間

（目的と到達目標）

相談者が相談者自身の言葉で疑問を入力した際に、求める情報に適切にナビゲートするチャットボットを試作し、将来の実装に必要な仕様を明らかにすることを目的とする。

- （１）日々の実践から FAQ データセットの蓄積方法の定式化
がん情報サービスサポートセンターで運営する電話相談、チャット相談の中で表出される疑問、質問とそれへの回答に用いられた情報について FAQ データセットについて、日々の業務の中で継続可能な形で蓄積する方法を明らかにする。
- （２）「情報提示チャットボット」のプロトタイプ作成、評価と活用可能シーンの抽出

領域を絞った範囲で、(1)で蓄積されたデータを基盤とするシナリオ型チャットボットと、AI型チャットボットを組み合わせた、相談者自身の言葉による疑問に対して情報提示を行う「情報提示チャットボット」のプロトタイプを作成する。

作成した「情報提示チャットボット」について、仮想ユーザーによる評価を行い、現実場面での活用可能性と、「有人チャット」や電話相談等との組み合わせでの利用等など、具体的な活用可能シーンを検討する。

- (3) がん情報の更新に対応できる、自動的な情報の蓄積方法を明らかにする
がん情報は常に更新されることから、提供する情報のアップデート（提示する URL やページ内の提示箇所の更新）を効果的に実施するための手法を明らかにする。

以上を通じて、情報提示チャットボットを実装し、継続的に運用する際に必要な手順と仕様を明らかにする。

当該年次

(到達目標)

(1) 日々の実践から FAQ データセットの蓄積方法の定式化

電話相談の自動音声認識データから効率的に FAQ 作成を行う方法を検討する

(2) 「情報提示チャットボット」のプロトタイプの作成、評価と活用可能シーンの抽出

領域を絞った範囲で、(1)で蓄積されたデータを基盤とするシナリオ型チャットボットと、AI型チャットボットを組み合わせた、相談者自身の言葉による疑問に対して情報提示を行う「情報提示チャットボット」のプロトタイプ Ver.1 を作成する。

(年次評価時点の実績要点)

(1) 日々の実践から FAQ データセットの蓄積方法の定式化

3 種類の FAQ 作成方法を検討した。

A：乳がんの電話相談 102 相談の逐語録から 978 件の FAQ を作成

B：チャット相談 54 相談から Q および Q を確認する QQ を抽出

C：相談員による記録 20 相談から Q と A のセットと、それを整えたもの、抽象度を上げたものを作成
この 3 種の方法を比較し、FAQ の蓄積に効率的、合理的な方法を検討した。A については、音声データからの逐語録作成、逐語録を構成する文字量が膨大であることから、作業負荷が過大であり、継続的な作成は困難であると判断した。B については、まだ手順の定式化の途上であるものの、相談内容と回答の情報内容について十分に理解できる人であればある程度の訓練で対応可能となる目途がたった。C については、経験値で補いながら作成する必要があるため、ベテラン相談員のみが持つスキルを必要としており、量産が困難であるが、量を絞れば日常業務の中で蓄積していく方法も検討しようと考えられた。

(2) 「情報提示チャットボット」のプロトタイプの作成、評価と活用可能シーンの抽出

(1) で作成した FAQ のうち 1014 件を用いて、「がん情報ボット Ver.5.2」を作成した。

回答パターンは

- ① 質問の要約+URL (がん情報サービス、乳癌診療ガイドライン)
- ② 質問の要約+複数のオプションから回答を選択
- ③ 回答文+URL (がん情報サービス、乳癌診療ガイドライン)
- ④ 回答文のみ

の 4 種であり、国立がん研究センター「患者・市民パネル」による試用、試用によるさらなる Q の収集と、使用感を問うアンケートを実施する。

研究成果と考察

当該次時評価時点

(1) 日々の実践から FAQ データセットの蓄積方法の定式化

FAQ データセットの効率的な蓄積方法を定式化するため、A 実際の電話相談データ、B 実際のチャット相談データ、C 相談対応を行っている相談員の経験、それぞれに基づく FAQ データセットの作成を試行した。また、電話相談データの音声認識による自動書き起こしの認識率向上に向けたデータ収集・分析を行った。

【A 実際の電話相談データに基づく FAQ データセットの作成】

2017 年 4 月 1 日～2019 年 6 月末の乳がんに関する相談の音声データ 102 件をテキスト化し、相談者の質問に基づいて簡潔な質問文 (Q) を作成した。Q の回答 (A) にあたる情報は、「がん情報サービス」または「患者さんのための乳がん診療ガイドライン」から探索し、該当する URL と組み合わせ FAQ を作成した。なお、Q に適する A が「がん情報サービス」「患者さんのための乳がん診療ガイドライン」にない場合は、相談員が対応する電話相談またはチャット相談を紹介する URL を提示することとした。URL の提示にあたっては、チャットボットが回答することを想定し、FAQ の内容に応じて定型回答文を作成した。定型回答文は以下 5 パターンとし、FAQ の内容に応じて、これらを組み合わせたデータセットを作成した。

①A で回答となる情報の URL を提示できない場合：

大変申し訳ございませんが、こちらでお答えするのは難しそうです。恐れ入りますが「がん電話相談」や「がんチャット相談」をご利用ください。相談員が直接お話を伺います。

②A で回答となる情報の URL を提示できる場合：

そのことでしたら、こちらの情報をご確認ください。ご参考になりましたら幸いです。

③A の提示のためにさらに具体的な Q の入力が必要な場合：

よろしければ、お知りになりたいことをもう少し具体的に

④不安や心配など気持ちに関する Q が入力された場合：

ご不安、ご心配があるのですね。このチャットボットでは十分にお力になれそうにありません。恐れ入りますが、「がん電話相談」や「がんチャット相談」をご利用ください。

⑤病院に関する情報を探す Q が入力された場合：

こちらのサイトから、国から指定を受けてがん医療を提供する全国の病院（がん診療連携拠点病院）を検索できます。より詳しい情報をお知りになりたい場合は、「がん電話相談」や「がんチャット相談」もご利用ください。

作成した FAQ は 978 件で、すべてのデータを「情報提示チャットボット」プロトタイプ作成の際の学習データとして用いた。

【B 実際のチャット相談データに基づく FAQ データセットの作成】

2020 年 11 月～2021 年 1 月のチャット相談テキストデータ 54 件、1178 発話（内訳：相談者 566 発話、相談員 612 発話）を用いた。相談者の発話から疑問や情報ニーズを含む可能性がある発話を抽出し、Q を作成した。相談員の発話からは、相談者の疑問を要約して明確化した発話、相談者の疑問を解釈した発話、相談者の疑問を明確にするための内容の確認にあたる発話を抽出し、相談員による相談者の疑問の要約・解釈に基づく質問 (QQ) を作成した。Q または QQ に対応する A に該当する記述を相談員の発話から抽出し、Q および QQ と組み合わせ FAQ を作成する方法を試行中である。

【C 相談員の経験に基づく FAQ データセットの作成】

2022 年 1 月 4 日～2022 年 1 月 20 日の 20 件の電話相談記録（相談毎に相談員が作成する電話相談の概要で、相談者のがん種や相談内容の概要が記載されている）を用いた。実際に日常的に電話相談にあたる相談員が、話相談記録から質問と回答にあたるデータをペアで抽出した FAQ、相談記録の文章を整えたり状況を補足したりするなどした FAQ、さらに抽象度を高めた相談記録に関連するよくある一般的な FAQ を作成した。作成した FAQ は 80 件で、このうち 36 件を「情報提示チャットボット」プロトタイプ作成の際の学習データとして用いた。

【電話相談データの音声認識による自動書き起こしの認識率向上に向けたデータ収集・分析】

2019～2020 年のがんと性に関する相談 22 件について、昨年度までの先行研究班での成果である改良した

「AmiVoice」を用いてテキストデータ化した。その結果、がんと性に関する相談は、がんや治療に関する相談よりも認識率が低かった。現在は、テキストデータ化した電話相談データと音声データを突合し、誤認識単語の抽出、誤認識の特徴的なパターンの有無や内容について分析・検討している。

以上より、電話相談データおよび相談員の経験に基づく FAQ データセットの作成については、素地が整いつつあると考えられる。しかし FAQ 作成においては、質問者の質問を正しく要約すること、正確な医学的知識に基づいて Q に対する A を検索し組み合わせることなど、一定の知識と能力が求められ、膨大な時間を要するのが現状である。がん専門相談員はこれらの能力を有しており、日々の実践の中で FAQ を作成することができると考えられる。今後、日々の相談実践と並行して FAQ データセットを作成する効率的な方法を検討していく。チャット相談データについては、相談時からテキスト化されたデータが存在することが強みであるものの、相談者のタイプミスや思い込みによる専門用語の誤用、相談の意図を把握するまでの相談者と相談員のやり取りをどのように FAQ データセットに落とし込むかなど、電話相談や相談員の経験に基づく FAQ 作成とは異なる課題があり、現在も検討を進めているところである。

今後は、(2)に記載した「情報提示チャットボット」プロトタイプの実行およびアンケート調査の結果をふまえて、①有人チャット相談の相談前振り分け機能の実装、②「情報提示チャットボット」推奨となる相談者へのチャットボットの返答形式を想定した対応パターンでの有人対応、③試行したデータを学習データとする蓄積形態の実装の検討を進めていく。また音声認識については、分析結果に基づいて音声認識ソフトのチューニング等を実施し、自動文字認識の精度向上に向けた取り組みを継続する。

(2)「情報提示チャットボット」のプロトタイプの実行、評価と活用可能シーンの抽出

相談者の質問に対して適した情報を提供するがん相談チャットボットの開発を目的に、乳がんに関する質問に回答する「情報提示チャットボット」のプロトタイプを作成した。想定ユーザーは乳がん患者とその家族で、(1)で作成した FAQ データ 1014 件 (A 実際の電話相談データに基づく FAQ 978 件、C 相談員の経験に基づく FAQ 36 件) を学習データとして用いた。チャットボットの型式は、チャットボットが誤った情報に利用者を誘導することや、情報を得た利用者が混乱することを避けるため、類似する質問の回答を表示するシナリオ型とした。

「情報提示チャットボット」のプロトタイプでは、質問者の入力に対するチャットボットの回答を①～④の4パターン作成した。

①質問の要約+情報 URL の提示

利用者が質問を入力すると、チャットボットが質問の要約と回答の候補となる URL を 3～5 つ提示する形式である (図 1)。提示する情報 URL は、「がん情報サービス」または「患者さんのための乳がん診療ガイドライン」とした。なお、質問に回答できない場合は個別相談の案内を表示する形式とした。

回答パターン①：質問の要約+URL



図 1：回答パターン①

②質問の要約+複数のオプションの提示

利用者が質問を入力すると、チャットボットが質問の要約とそれに関連する複数のオプションを提示する。利用者は、チャットボットが要約した質問を確認し、要約した質問に対して提示される複数のオプションからあてはまるものを全て選択し、用者の選択に応じた回答や回答に関連する情報 URL を提示する形式である（図 2）。提示する URL は、「がん情報サービス」または「患者さんのための乳がん診療ガイドライン」とした。

回答パターン②：質問の要約+複数のオプションから回答を選択

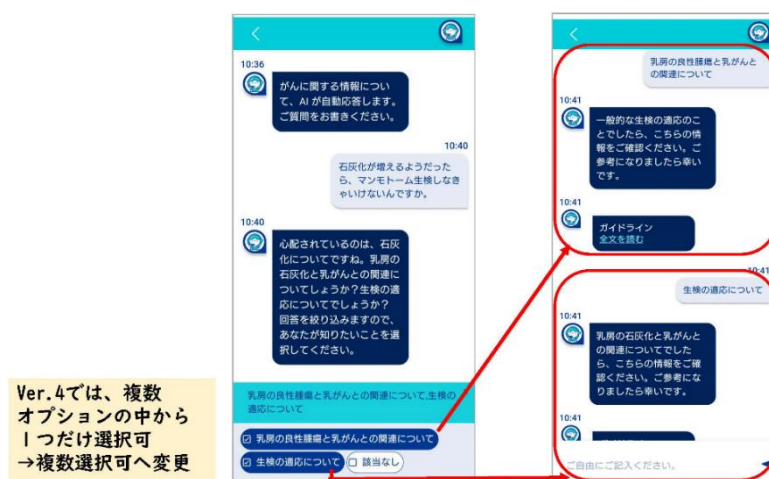


図 2：回答パターン②

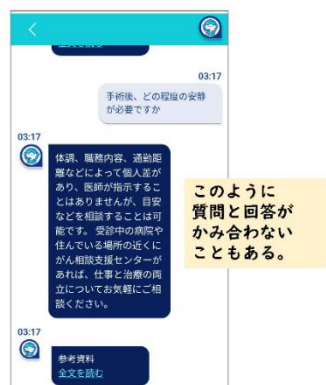
③回答文+情報 URL の提示

利用者の質問に対し、チャットボットが質問の要約を提示することなく回答文と回答に関連する URL を提示する形式である（図 3）。提示する URL は、「がん情報サービス」または「患者さんのための乳がん診療ガイドライン」とした。

④回答文のみ提示

利用者の質問に対し、チャットボットが質問の要約や情報 URL を提示せず、回答文のみを提示する形式である（図 3）。

回答パターン③：回答文+URL



回答パターン④：回答文のみ

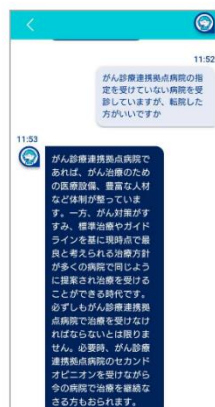


図 3：回答パターン③④

(3) がん情報の更新に対応できる、自動的な情報の蓄積方法を明らかにする
上記(1)(2)におけるFAQデータセット作成の際に、がん情報サービスおよび外部サイトにおける情報のアップデートへの対応方法を検討しているところであり、今後のFAQ更新手順の検討の中で進めていく

(1) 遵守すべき研究に関する指針等

- ☐ 再生医療等の安全性の確保等に関する法律
- ☒ 人を対象とする生命科学・医学系研究に関する倫理指針
- ☐ 遺伝子治療等臨床研究に関する指針
- ☐ 動物実験等の実施に関する基本指針
- ☐ 臨床研究法
- ☐ 医薬品の臨床試験の実施に関する省令（GCP）
- ☐ その他の指針等（指針等の名称：_____）

(2) 本研究開発期間中に予定される臨床研究の有無

- ☐ 有
- ☒ 無

チャットボットのプロトタイプへの評価調査については、国立がん研究センター研究倫理審査委員会の審査を受け、実施した(2023-270)。

当該年次

- ・国立がん研究センター研究開発費による成果であることが記載されているもの
記載なし

- ・国立がん研究センター研究開発費による成果であることが記載はないが、関連するもの記載なし

・堀抜文香, 稲垣安沙, 櫻井雅代, 志賀久美子, 宮本紗代, 八巻知香子. がんチャット相談の特徴と相談員が活用するスキル: がん専門相談員へのフォーカスグループインタビューにもとづいて. 第8回日本がんサポーターティブケア学会学術集会, 2023.6.22. 奈良.

- 7-

(書籍)

記載なし

(知的財産権)

記載なし

(政策提言（寄与した指針等）)

記載なし

(その他)

記載なし