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）

AIによる参照利用を想定したがん診療ガイドラインの標準的Web構造の有用性

研究協力者　西迫　宗大　国立がん研究センター がん対策研究所 がん情報提供部（特任研究員）
研究協力者　Mingxin Liu　国立がん研究センター がん対策研究所 がん情報提供部（特任研究員）
研究分担者　東 尚弘　東京大学大学院 医学系研究科 公衆衛生学分野（教授）
研究代表者　若尾 文彦　国立がん研究センター がん対策情報センター本部（副本部長）

研究要旨

近年、検索拡張生成（Retrieval-Augmented Generation：RAG）を用いた生成 AI は、医療情報提供手段として注目されている。特に診療ガイドラインは、AI が参照する信頼性の高い情報源として期待されるが、現在の Web ガイドラインは学会・団体ごとに異なる構造で運用されており、AI による情報抽出効率や回答品質への影響は十分検討されていない。本研究では、AI による診療ガイドライン活用を見据え、Web ガイドライン構造と RAG 型 AI チャットボットの応答性能との関連を検討した。研究対象として、Web 上で無料公開されている 13 編のがん関連診療ガイドラインおよび患者向けガイドブックを解析した。各ガイドラインについて、目次ページを対象に Python（BeautifulSoup ライブラリ）によるスクレイピング処理を行い、抽出された URL の内容を解析した。さらに、Web 構造を「Clinical Question（CQ）単位・個別指定」型、「節単位・アンカー指定」型、「PDF 全文・スクロール」型、「章単位・スクロール」型、「別サイト・個別指定」型の 5 種類に分類した。加えて、ガイドラインを参照して回答を生成する RAG 型 AI チャットボットを構築し、各構造における情報抽出効率および応答性能を比較した。その結果、CQ ごとに独立した HTML ページと URL を有する「CQ 単位・個別指定」型では、抽出 URL の大部分が医療情報を含み（104/106、98%）、AI は高精度に関連情報を抽出し、ガイドライン内容に整合した回答を生成可能であった。一方、章単位構造や全文 PDF 型では、重複情報の増加や LLM 入力制限の影響により、情報抽出効率および回答性能の低下が認められた。また、参照情報が不足する場合には、「情報不足」等の応答を生成し、不適切な回答を抑制できることを確認した。本研究により、AI を活用した診療ガイドライン情報提供においては、CQ 単位で独立したページ構造を有する Web 設計が重要であることが示された。今後、AI 利用を前提とした Web ガイドライン構造の標準化が進めば、複数ガイドラインを横断的に参照する汎用的な医療情報提供 AI の実装が可能となり、安全かつ信頼性の高い医療情報基盤の構築につながると考えられる。

A．研究目的

近年、インターネットおよびソーシャルメディアの発達に伴い、患者や医療従事者がWeb上から医療情報を取得し、診療や意思決定に活用する機会は大きく増加している。特にがん医療領域では、診療ガイドラインが標準治療や診療方針を示す重要な情報基盤として広く利用されている。一方で、ガイドラインの内容は年々高度化・複雑化しており、必要な情報へ迅速に到達することは容易ではない。また、Web上には信頼性の低い医療情報も存在することから、正確かつ根拠に基づく情報提供の重要性が高まっている。

近年では、大規模言語モデル（Large Language Model：LLM）を用いた生成AIが急速に発展し、自然言語による医療情報検索や要約への応用が期待されている。しかし、生成AIには事実に基づかない情報を生成する hallucination の問題が存在し、医療分野においては安全性および信頼性の観点から大きな課題となっている。これに対し、外部情報を参照しながら回答を生成する Retrieval-Augmented Generation（RAG）は、根拠に基づく情報提示を可能とする技術として注目されている。特にWeb上で公開された診療ガイドラインは、将来的にAIが参照する主要な情報源となることが想定される。

一方で、現在の診療ガイドライン作成指針においては、AIによる活用を前提としたWeb構造設計について十分な検討がなされていない。そのため、Web構造の違いがAIによる情報検索効率や回答品質にどのような影響を与えるかについては、いまだ明らかではなく、その有用性に関する体系的な検討も十分に行われていない。このことは、AIが必要な情報へ適切

にアクセスできない可能性を生じさせるだけでなく、情報抽出漏れや不適切な回答生成、さらにはhallucination発生の一因となる可能性がある。特に医療分野では、回答の不正確性が患者や医療従事者の意思決定へ直接影響を与える可能性があることから、AIが参照しやすい情報構造設計の重要性は今後さらに高まると考えられる。

本研究では、Web上で公開されている複数のがん診療ガイドラインおよび患者向けガイドブックを対象として、Web構造の分類および情報抽出効率の評価を行った。さらに、診療ガイドラインを参照するRAG型AIチャットボットを構築し、各Web構造がAIによる情報検索および回答生成に与える影響を解析した。これにより、将来的なAI活用を見据えた「AIが読み取りやすい診療ガイドライン構造」の要件を明らかにし、医療分野における安全かつ信頼性の高い情報提供基盤の構築に資することを目的とした。

B. 研究方法

解析対象

事前に研究利用に関する許可を取得した、Web 上で無料公開されている 13 編のがん関連診療ガイドラインおよび患者向けガイドブックを解析対象とした（表 1）。対象には、肺がん、乳がん、GIST、甲状腺腫瘍、脳腫瘍等の個別がん診療ガイドラインに加え、高齢者薬物療法、リハビリテーション、G-CSF 適正使用、リンパ浮腫、妊孕性温存等の横断的ガイドライン、および患者向けガイドブックを含めた。

具体的には、「患者さんと家族のための肺がんガイドブック 2024 年版」（以下、肺がん患者）、「肺癌診療ガイドライン―悪性胸膜中皮腫・胸腺腫瘍含む 2024 年版」（肺癌診療）、「GIST 診療ガイドライン 2022 年 4 月改訂 第 4 版」（GIST）、「乳癌診療ガイドライン 2022 年版 治療編総説・治療編」（乳癌診療）等を対象とした。

対象ガイドラインについて、目次ページ URL およびガイドライン本文へのリンク構造を確認した（図 1）。

サイト構造の把握と分類

各ガイドラインについて、ホームページURL、目次ページURL、Webサイト運営主体、公開形式（HTML／PDF）を記録した。さらに、ガイドラインページへアクセスし、HTMLソースおよびディレクトリ構造を確認することで、ガイドライン本文コンテンツの配置形式を解析した。解析では、（1）情報提示単位（CQ単位、節単位、章単位、全文PDF等）、（2）コンテンツ指定方法（個別リンク、アンカーリンク等）、（3）ファイル形式（HTML／PDF）の観点から、Web構造の分類を行った。

WebスクレイピングによるAI参照情報抽出

各ガイドラインの目次ページを対象として、Python（version 3.12.12）およびBeautifulSoupライブラリを用いたスクレイピング処理を実施した。ページ内から抽出された全URLを記録し、それぞれについて手動で内容を確認した。抽出されたURLは、Clinical Question（CQ）、Background Question、解説本文等の「医療情報を含むページ」と、学会トップページや案内ページ等の「医療情報を含まないページ」に分類した。また、同一または類似した内容を示すURLについては「重複ページ」として記録した。

各ガイドラインについて、抽出された全URL数に対する医療情報ページ数および重複ページ数の割合を算出し、Web構造ごとの情報抽出効率を比較した。

ガイドライン参照型RAGチャットボットの構築

診療ガイドラインを参照して回答を生成するRAG型AIチャットボットを構築した。システムは、（1）質問文からの検索キーワード生成、（2）類似度に基づく関連ページ抽出、（3）参照情報生成、（4）LLMによる回答生成、の4段階で構成した。

ユーザー質問から生成された検索キーワードと、各ガイドラインページの本文をMeCab（version 0.996）により形態素解析した後、TF-IDFによりベクトル化した。コサイン類似度を算出し、一定以上の類似度（0.2以上）を示したページを参照情報として抽出した。抽出されたページ情報を統合した後、API経由 (OpenAI, Inc., San Francisco, CA, USA) でLLM（GPT-3.5-turbo）へ入力し、ガイドライン内容に基づく回答を生成した（図2）。

RAG型AIチャットボット応答性能の評価

胸腺腫瘍、甲状腺腫瘍、GIST、非浸潤性乳管癌等を対象として、診断法、病理診断、治療法、外科治療等に関する質問文を作成した。各質問に対して、RAG型AIチャットボットが参照したページ数、実際に有効であった参照情報数、コサイン類似度、回答生成の有無、および回答内容を記録した。さらに、参照ガイドラインに含まれない疾患に関する質問を入力し、AIが不適切な回答を生成することなく、「情報不足」等の応答制御を適切に行えるかを検証した。

（倫理面への配慮）
本研究は公開情報のみを対象としており、個人情報を取り扱わないため、倫理上の問題は生じないと考えられる。

C. 研究結果

1. Webガイドライン構造の分類

解析対象とした13編のWebガイドラインは、Web構造の特徴に基づき5種類に分類された（表1）。

CQごとに独立したHTMLページを有し、個別URLによって構成される「CQ単位・個別指定」型には、「肺がん患者」「乳がん患者」「遺伝性卵巣」が含まれた。

単一HTMLページ内で章構造を持ち、アンカーリン

クによって各CQを指定する「節単位・アンカー指定」型には、「肺癌診療」が分類された。本構造では、URLによる指定は可能であるものの、実際には同一章全体を参照する形式であり、CQ単位で独立したページ構造は認めなかった。

ガイドライン全体が単一PDFファイルとして提供される「PDF全文・スクロール」型には、「甲状腺腫瘍」や「リハビリ」が含まれた。本構造では、ガイドライン全体が一括して提供されスクロール操作によって閲覧する形式であった。

また、章単位で長文HTMLページが構成される「章単位・スクロール」型、および目次ページから別ディレクトリや外部ファイルへリンクされる「別サイト・個別指定」型が認められた。

2. 目次ページからの医療情報抽出効率

各ガイドラインの目次ページを対象としてスクレイピングを行い、抽出されたURLのうち、実際に医療情報を含むページの割合を比較した（図3）。

CQ単位・個別指定型である「肺がん患者」では、106件のURLが抽出され、そのうち104件（98%）が医療情報を含むページであった。重複ページは認めず、CQ単位で高効率な情報抽出が可能であった（表2）。

節単位・アンカー指定型である「肺癌診療」では、384件のURLが抽出され、そのうち380件（99%）が医療情報を含んでいた。しかし、異なるURLが同一章を参照する構造であったため、実質的には重複した情報抽出が多数認められた。

PDF全文・スクロール型では、ガイドライン全体が単一PDFとして提供されていたため、抽出可能なURLは1件のみであり、CQ単位での情報抽出は不可能であった。

章単位・スクロール型では、多数のURLが抽出されたものの、医療情報を含む割合は低く、対象外ページや補助情報ページが多く含まれていた。また、別サイト・個別指定型では、目次ページから直接ガイドライン本文へアクセスできない構造を認め、ガイドライン本体コンテンツの効率的抽出は困難であった。

3. RAG型AIチャットボットの応答性能

構築したRAG型AIチャットボットを用いて、診断法、病理診断、治療法、外科治療に関する質問を入力し、Web構造ごとの応答性能を比較した（表2）。

CQ単位・個別指定型では、全ての質問に対して適切な回答生成が可能であった。

一方、節単位・アンカー指定型では、多数のページが参照情報として抽出されたものの、同一章を重複して参照する傾向を認めた。その結果、関連情報量は多いものの、有効な参照情報の割合は低下し、一部の質問では適切な回答生成ができなかった。

PDF全文・スクロール型では、ガイドライン全文が参照対象となるため、LLMの入力制限内では文書冒頭部分のみが参照情報として利用された。その結果、質問内容に関連する情報を参照することができず、回答生成が困難であった。

「章単位・スクロール」型および「別サイト・個別指定」型でも関連情報の不足のため回答生成ができなかった（データ示さず）。

4. 参照情報不足時における応答制御

参照情報が十分に存在しない場合、RAG型AIチャットボットは「テキスト情報には十分な情報がありません」と回答を生成し、不適切な推測回答を抑制した（表4）。さらに、肺がんガイドラインを参照情報とした状態で甲状腺に関する質問を入力するなど、参照範囲外の質問を行った場合においても、RAG型AIチャットボットは意味的関連性の低さを認識し、回答生成を抑制した（図4）。

5. AI利用を想定した標準的Web構造の検討

本研究において最も高効率に医療情報を抽出できたのは「CQ単位・個別指定」型であった。特に、「肺がん患者」では、CQごとに独立したURLが付与されており、共通ヘッダーや不要な外部リンクを最小限に抑えたシンプルな構造を有していた。

以上の結果を踏まえ、AIによる診療ガイドライン活用を想定した場合には、CQ単位で独立したページ構造を有し、個別URLによって直接参照可能なWeb構造が最も適していると考えられた（図5）。

D. 考察

本研究では、Web上で公開されている複数のがん診療ガイドラインを対象としてWeb構造を分類し、それぞれの構造がRAG型AIチャットボットによる情報検索および回答生成に与える影響を検討した。その結果、CQごとに独立したHTMLページを有する「CQ単位・個別指定」型では、AIが高効率かつ高精度に関連情報を抽出でき、ガイドライン内容に整合した回答生成が可能であることが示された。

本研究で用いたスクレイピング手法およびRAG構成は、Pythonによる比較的単純なWebスクレイピングと、TF-IDFおよびコサイン類似度を用いた基本的な情報検索アルゴリズムによって構築されたものであった。それにもかかわらず、「CQ単位・個別指定」型では高い応答性能が得られたことから、AI活用において重要なのは複雑なAIアルゴリズムそのものではなく、「AIが利用しやすいWeb構造」が整備されているかどうかである可能性が示された。特に、CQ単位で独立したURLが付与された構造では、AIが必要な情報へ直接到達できるため、不要情報の混入や重複参照を低減することが可能であった。一方、「節単位・アンカー指定」型では、異なるURLが同一章を参照することで重複情報が多数生成され、参照効率の低下が認められた。また、「PDF全文・スクロール」

型では、LLMの入力可能文字数制限により文書全体を十分に参照できず、質問内容との関連性が低い情報のみがAIへ入力される結果となった。さらに、本研究では、参照情報が存在しない場合や、参照範囲外の質問が入力された場合に、RAG型AIチャットボットが「十分な情報が存在しない」と応答し、不適切な回答生成を抑制できることを確認した。この結果は、RAG構造がhallucination抑制に有効である可能性を示唆しており、医療分野における安全性の高いAI情報提供システム構築につながる知見であると考えられた。

各ガイドラインのユーザーインターフェイス（UI）とAIによる情報抽出効率を比較すると、「人にとって閲覧しやすい構造」と「AIにとって処理しやすい構造」は必ずしも一致しないことが明らかとなった。人間にとっては、階層的メニューや折りたたみ表示、統合的なナビゲーションを備えたUIは直感的で理解しやすい。一方、AIにとっては、CQ単位で分割されたHTML構造、重複の少ない独立URL、不要なヘッダー・フッターを含まないシンプルな構造の方が情報抽出効率に優れていた。今後、AIを利用した診療支援や患者向け情報提供を推進するためには、人間およびAI双方にとって利用しやすいWebガイドライン構造を検討していく必要がある。

現在のWebガイドラインは、学会・団体ごとに構造が異なっており、AI利用を前提とした標準化は行われていない。そのため、AIによる情報抽出には個別対応が必要となり、汎用的な情報提供システム構築の障壁となっている。今後、AI活用を前提としたWebガイドライン構造の標準化が進めば、単一システムで複数ガイドラインを横断的に参照可能となり、患者および医療従事者に対する正確かつ効率的な医療情報提供につながる可能性がある。

E. 結論

本研究により、診療ガイドラインを AI が効率的かつ正確に利用するためには、CQ 単位で独立したページ構造を有する Web 設計が重要であることが示された。「CQ 単位・個別指定」型では高い情報抽出効率と応答性能が得られた一方、章単位構造や全文 PDF 型では、情報抽出効率および AI 応答性能の低下が認められた。また、RAG 型 AI チャットボットは、適切な参照情報が存在する場合にはガイドライン内容に基づく回答生成が可能であり、参照情報が不足する場合には回答を抑制することで、hallucination 抑制に寄与する可能性が示された。

以上より、今後のAI活用を見据えた診療ガイドライン整備においては、「AIが読み取りやすいWeb構造」を考慮した標準化が重要であると考えられる。Web構造の標準化が進めば、複数ガイドラインを横断的に参照可能な汎用的医療情報提供AIの実装が可能となり、患者および医療従事者に対する安全かつ信頼性の高い情報提供基盤の構築につながることが期待される。

F. 健康危険情報

特になし

G. 研究発表

1. 論文発表

Nishisako S, Higashi T, Wakao F. Design and Development of an AI Chatbot System Referencing Web-Based Cancer Clinical Practice Guidelines, *J Vis Exp.* under review.

2. 学会発表

西迫宗大、東尚弘、若尾文彦. Webガイドライン参照型AI-Chatbotの開発 ~標準的なサイト構造の有用性~ 第62回日本癌治療学会学術集会, 横浜市, 2025/10/16.

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

ガイドライン（Web版；13編）

略称	名称
肺がん患者	患者さんと家族のための肺がんガイドブック 2024年版
肺癌診療	肺癌診療ガイドライン-悪性胸膜中皮腫・胸腺腫瘍含む 2024年版
GIST	GIST診療ガイドライン 2022 年4月改訂 第4版
乳癌診療	乳癌診療ガイドライン 2022年版　治療編総説・治療編
甲状腺腫瘍	甲状腺腫瘍診療ガイドライン 2024
乳がん患者	患者さんのための乳がん診療ガイドライン 2023年版
脳腫瘍	脳腫瘍診療ガイドライン
遺伝性乳卵巣	遺伝性乳がん卵巣がんを知ろう！
高齢者薬物	高齢者のがん薬物療法ガイドライン
リハビリ	がんのリハビリテーション診療ガイドライン (第2版)
G-CSF	G-CSF適正使用ガイドライン 2022年10月改訂　第2版
リンパ浮腫	リンパ浮腫診療ガイドライン 2018年版 (第3版)
妊孕性	小児，思春期・若年がん患者の妊孕性温存に関する診療ガイドライン 2017 年版

図1. 本研究で用いたWeb ガイドライン

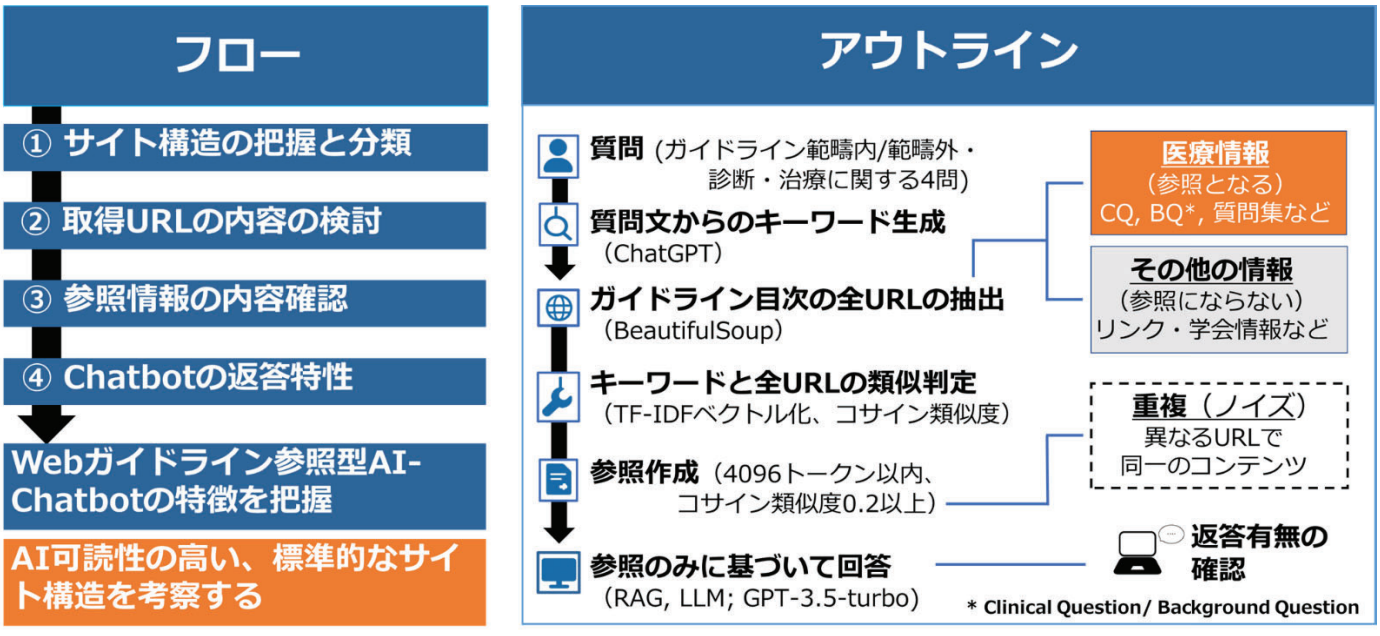


図2. 解析フロー

表 1. Web ガイドライン構造の分類

タイプ	収納場所	構造単位	CQの指定	形式	該当ガイドライン
① CQ単位・個別指定	目次フォルダ内	CQごと	ハイパーリンク	HTML	**肺がん患者***, 乳がん患者, 遺伝性乳卵巣
② 節単位・アンカー指定	目次フォルダ内	節ごと	アンカーリンク	HTML	**肺癌診療***
③ PDF全文・スクロール	目次フォルダ内	全文	なし	PDF	**甲状腺腫瘍***, リハビリ
④ 章単位・スクロール	目次フォルダ内	章ごと	なし	HTML	**GIST***, 高齢者薬物, G-CSF, リンパ浮腫, 妊孕性
⑤ 別サイト・個別指定	目次フォルダ外	CQごと	ハイパーリンク	HTML	**乳癌診療***, 脳腫瘍

*解析に使用

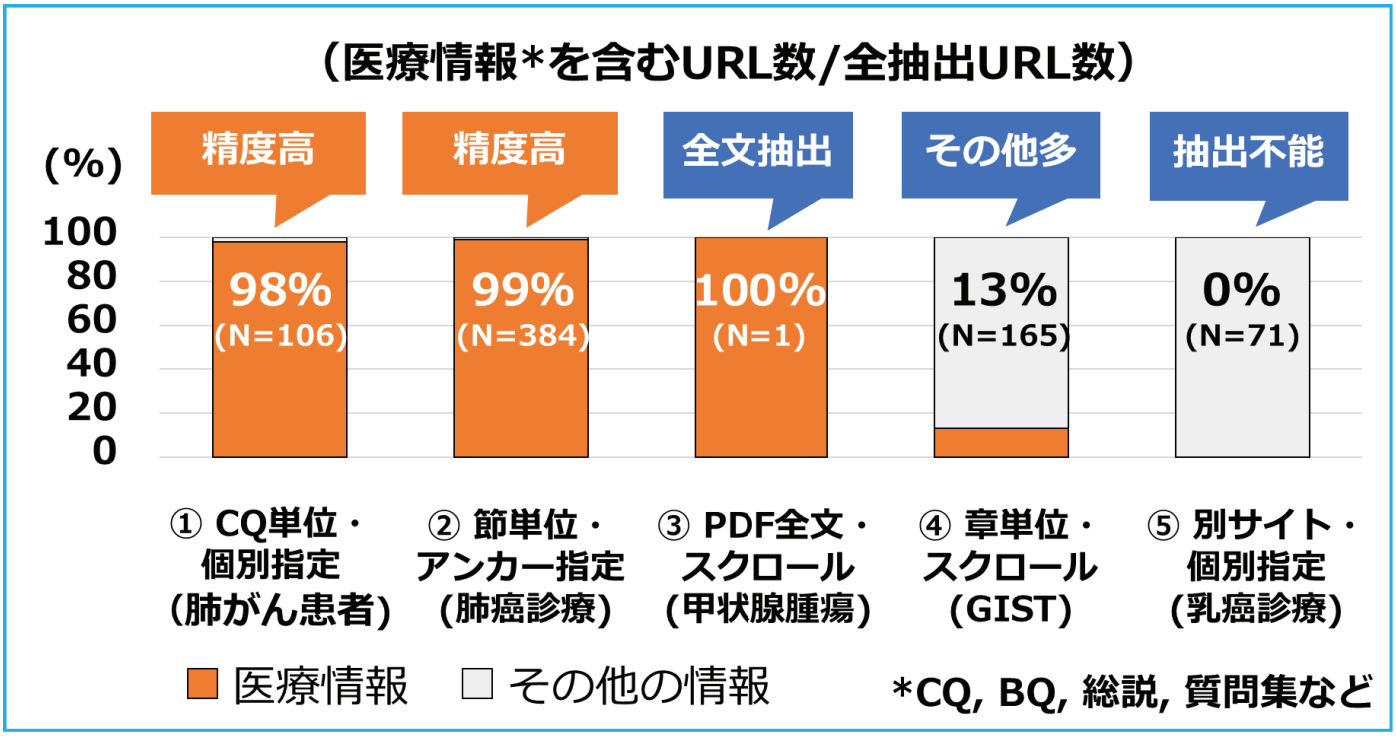


図3. 目次ページからの医療情報抽出効率

表. 2 ガイドライン構造によるRAG型AIチャットボットの応答性能

タイプ （ガイドライン, 全抽出数）	参照した項目数, *n*, range	†実質的参照項目, *n*, range (%)	Chatbot 返答有無
① CQ単位・個別指定 （患者向け肺がん, N = 106）	3-15	3-14 (93-100)	100% (4/4)
② 節単位・アンカー指定 (肺癌診療, N = 384)	40-75	12-18 (24-31)	75% (3/4)
③ PDF全文・スクロール (甲状腺腫瘍, N = 1)	1	‡11%	0% (0/4)

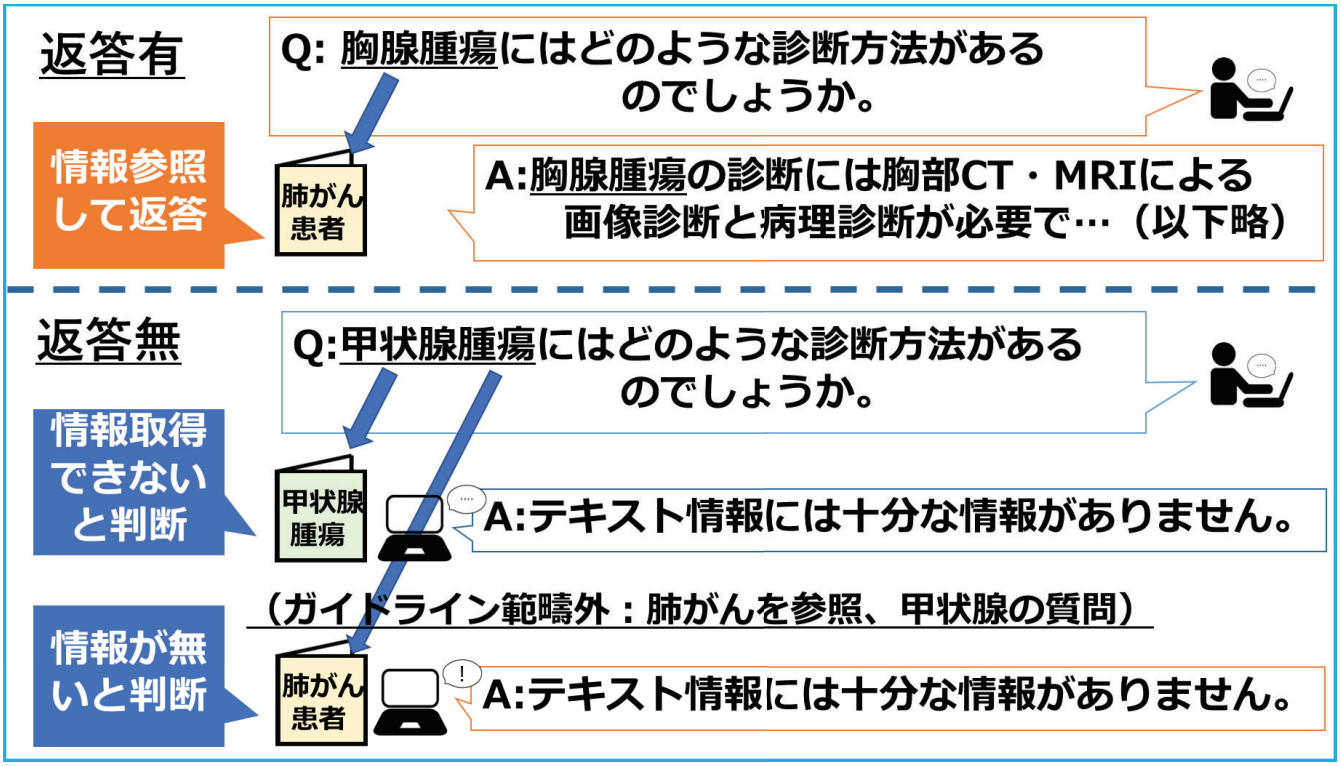


図4. RAG型AIチャットボットの応答例　（情報参照時および参照情報不足時）

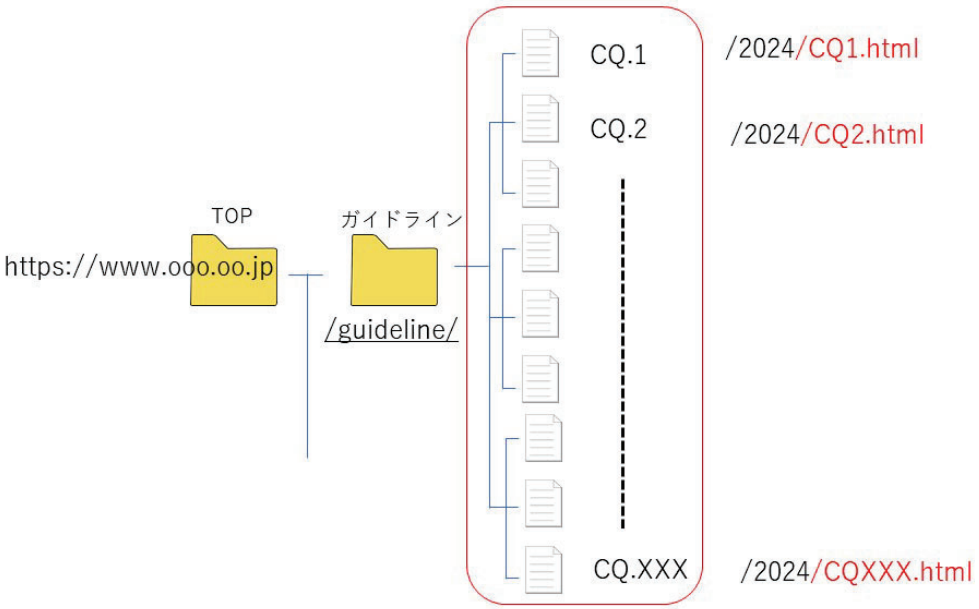


図5. AI活用を目的としたWebガイドラインの構造案