

臨床研究における生成AIの活用の可能性

大阪大学医学部附属病院
未来医療開発部 臨床研究センター
副センター長 浅野 健人

本発表には開示すべき利益相反事項はありません。

令和5年度厚生労働科学特別研究事業
(R6/1/16～R6/3/31)

生成AIを用いた治験・臨床研究関連文書の
デジタルトランスフォーメーション
に向けた研究

成果物は以下に公開

<https://mhlw-grants.niph.go.jp/project/167945>

本研究の目的

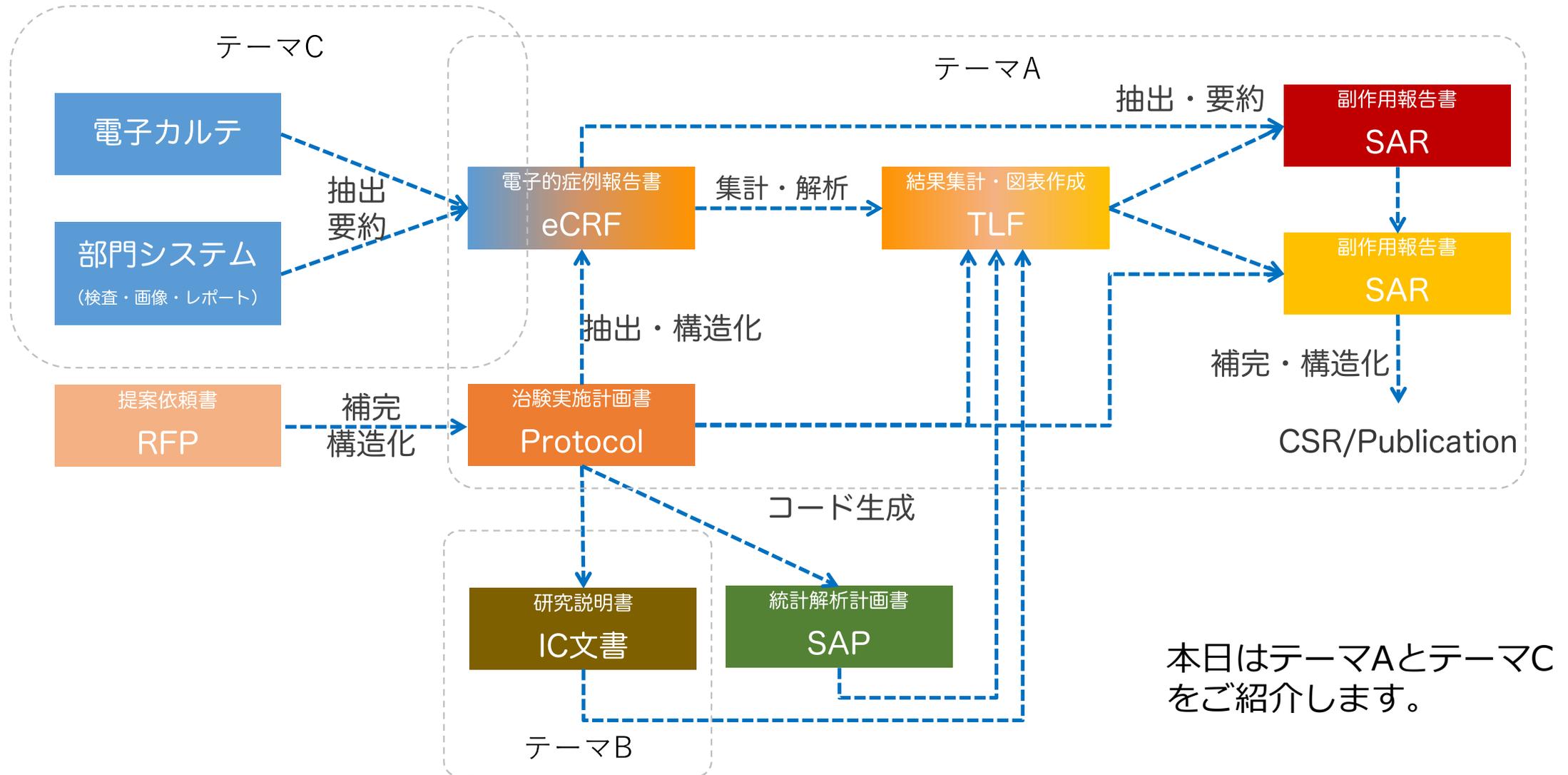
臨床研究の実施における関連文書にまつわる生成AIの活用について、完全自動化から人間と生成AIの協働まで、既存領域知識をどのように取り込み、精度を向上させながら安全かつ効果的に利用できるかを検証・実証することを目的とする。

具体的には、下記の課題、仮説を検討する。

- 臨床研究・治験は書類作成や事務作業に膨大な時間を取られる（課題）
- 必ずしも、人が実施しなくとも、昨今話題になっている生成AI等の技術を上手に活用することが出来ないか？（仮説）
 - AIプロダクト（DIP社 WritingX）に既存情報を活用すると、プロトコルの自動生成は出来るか？（テーマA）
 - 開発中のAI（サイバーエージェント社LLM）を活用することにより、患者が理解をするのに時間を要する説明文書を理解しやすいように平易化出来るか？（テーマB）
 - 電子カルテ出力情報をAIプロダクト（DIP社 WritingX）に引用させることにより、情報の抽出を行い、CRFの自動入力できるか？（テーマC）

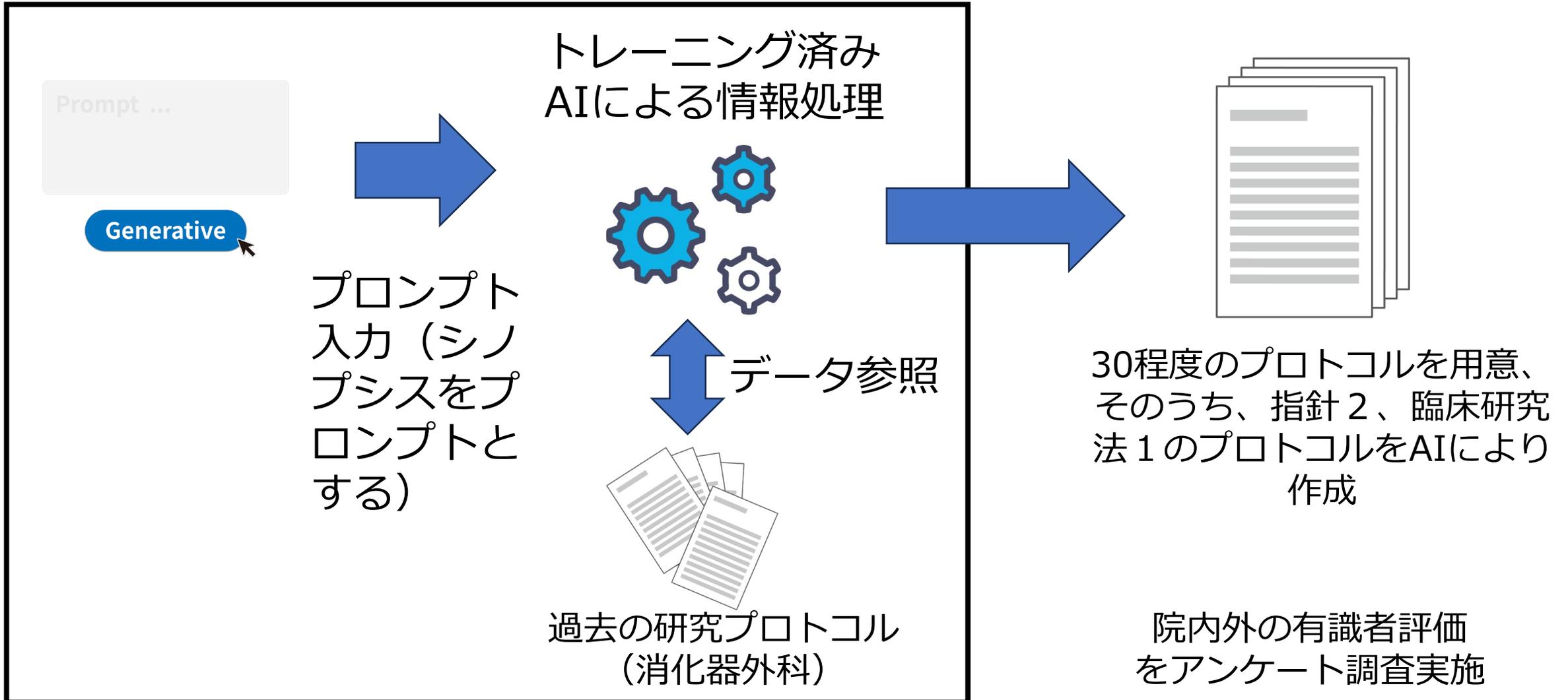
本研究の全体像

大規模言語モデルは治験文書の自動生成・臨床データ自動入力に有用か？



テーマA プロンプトによる計画書の自動作成

生成AIソフトウェア (DIP社)



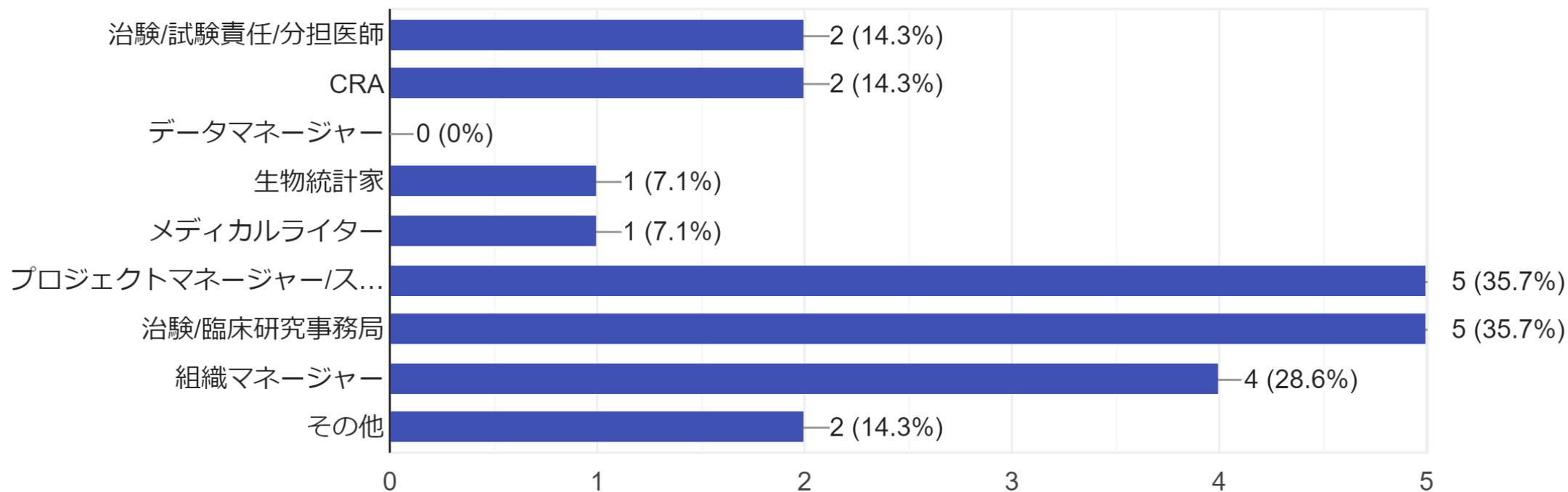
生成AIで作成されたプロトコルの出来上がりの前提と補足と所感

- 消化器外科領域の手術試験を中心に30程度のプロトコルを用意
- RAG (Retrieval-Augmented Generation) という技術を用いる
- ターゲットプロトコルを決定すると、目次を引用する作業を自動で行う
- プロンプト情報としては、A4 1ページ、1000字のプロトコルシノプシスの情報を提供する
- 特別な設定等を行わずに、A4 17ページ、2万字のプロトコルが出来上がった
- 繰り返し表現などが頻出して、読み物としては読みにくい箇所あり
- 項の間など、ところどころ矛盾する箇所もあり
- 背景などの箇所は決して十分な内容ではない
- 雛形の表現として、残しておいてほしい箇所を残してはくれない

アンケート

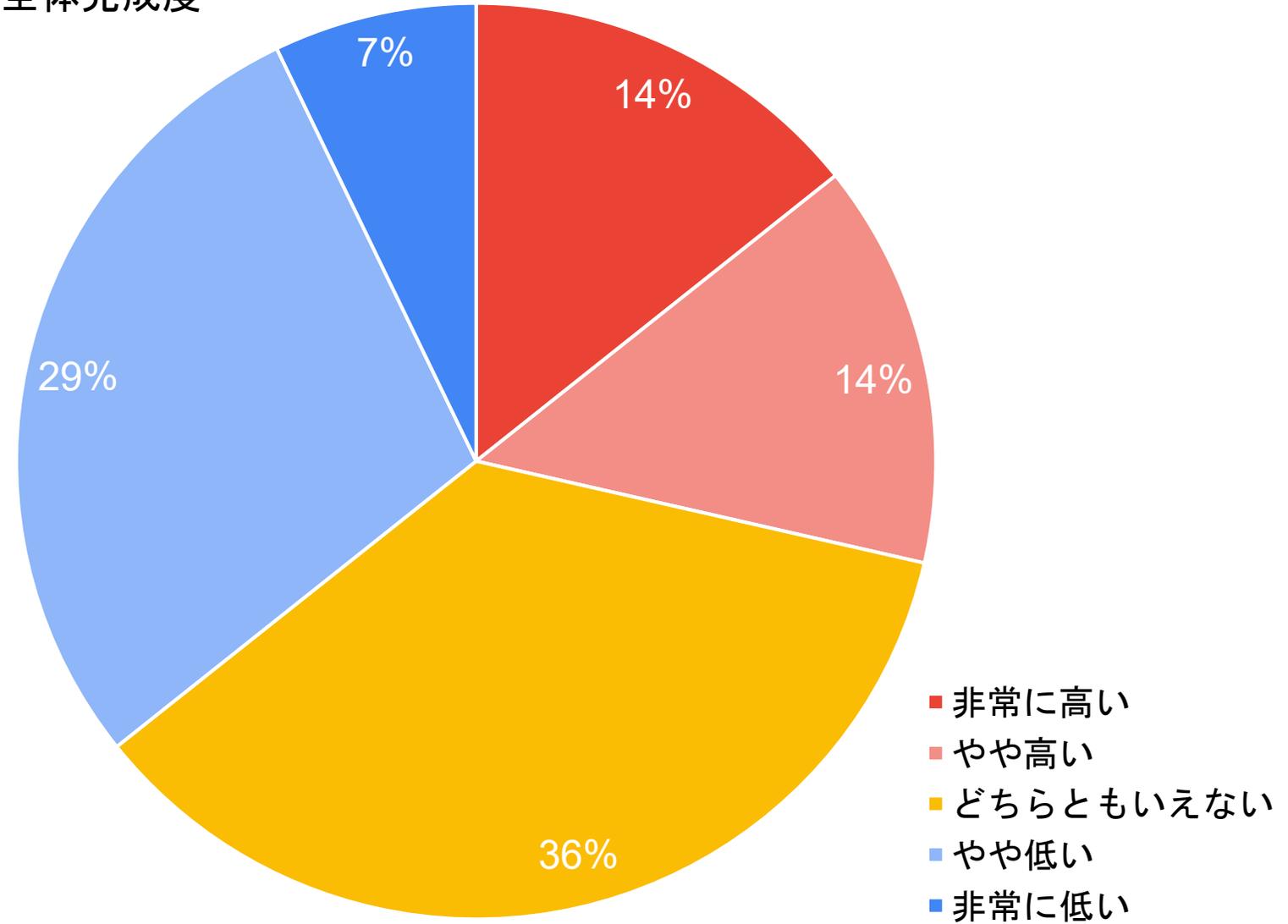
2. あなたの職種（主な役割）について ※複数選択可

14件の回答

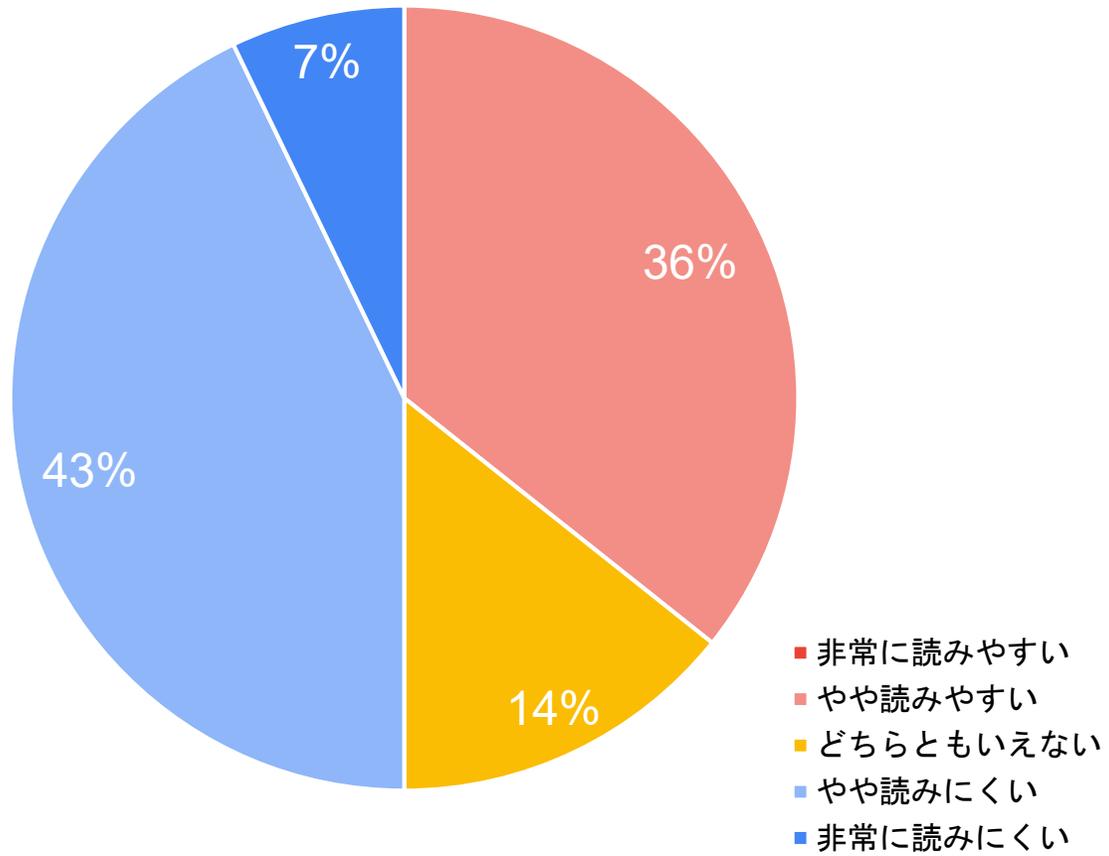


プロトコルの印象

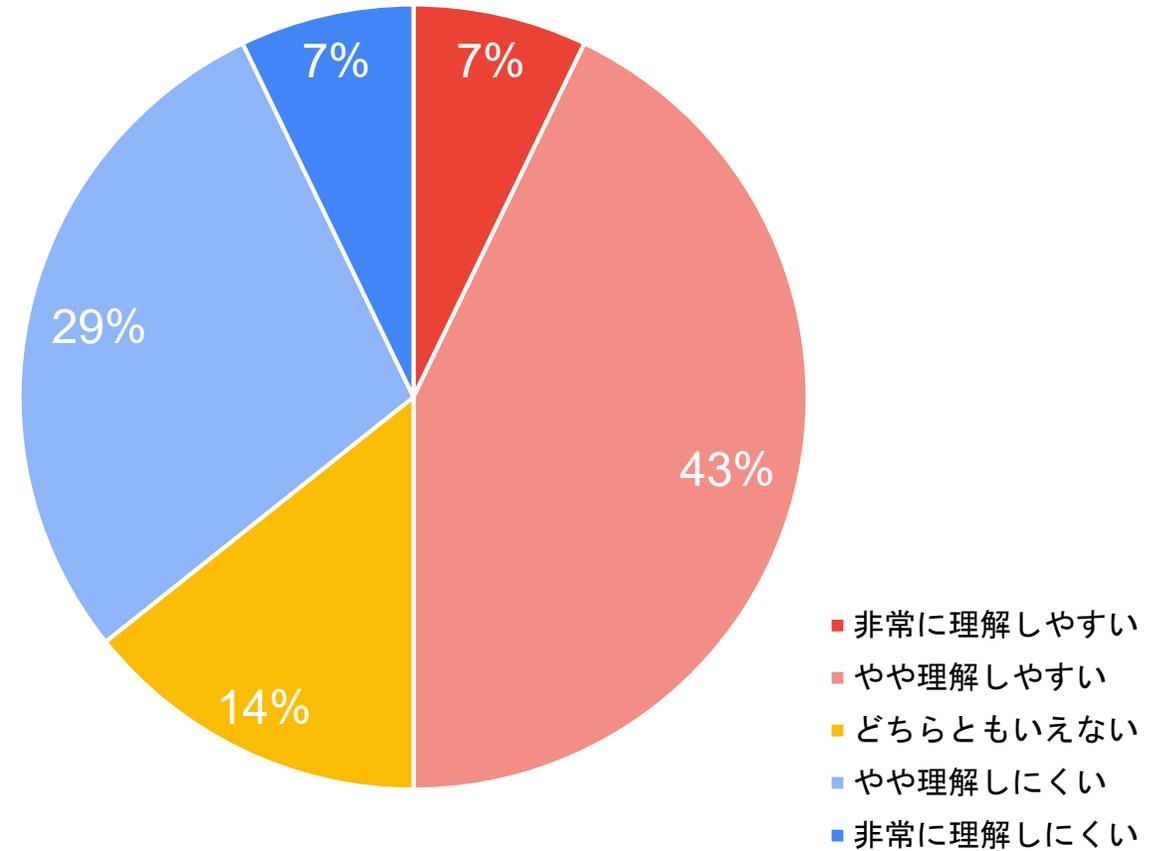
1. 全体完成度



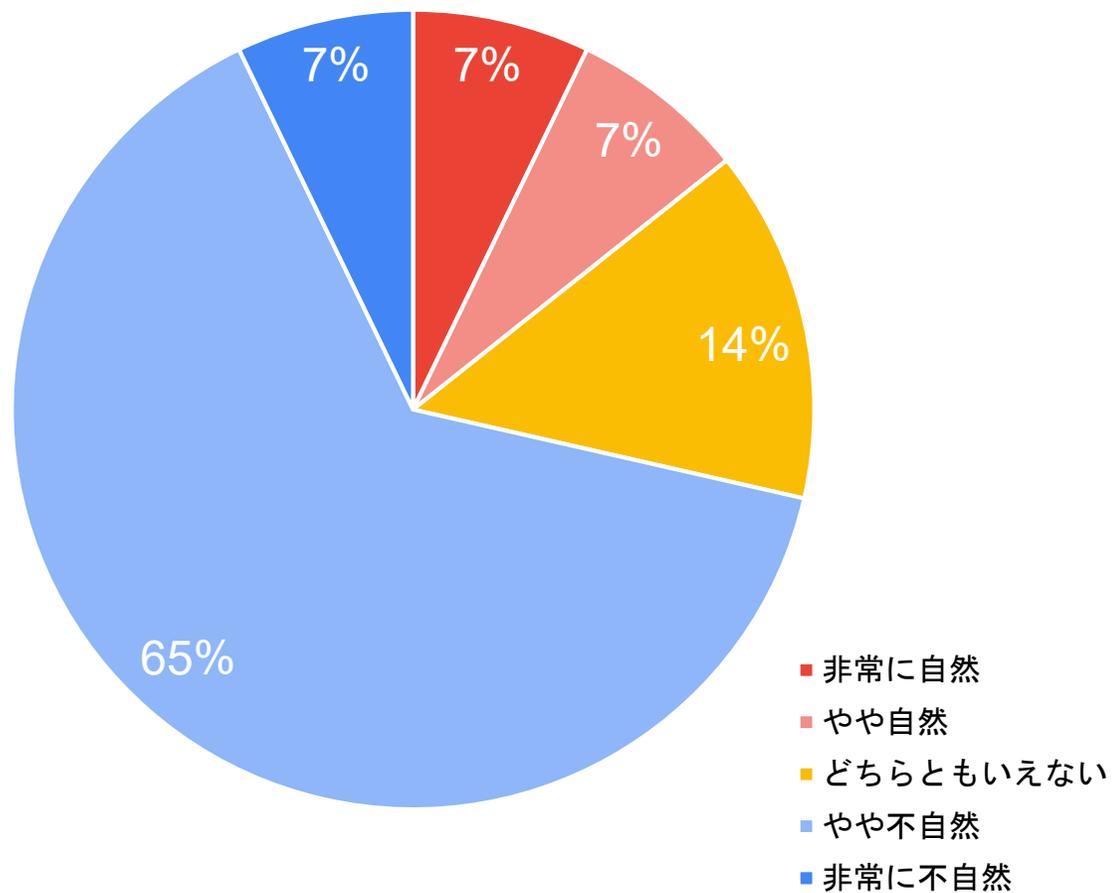
2. 読みやすさ



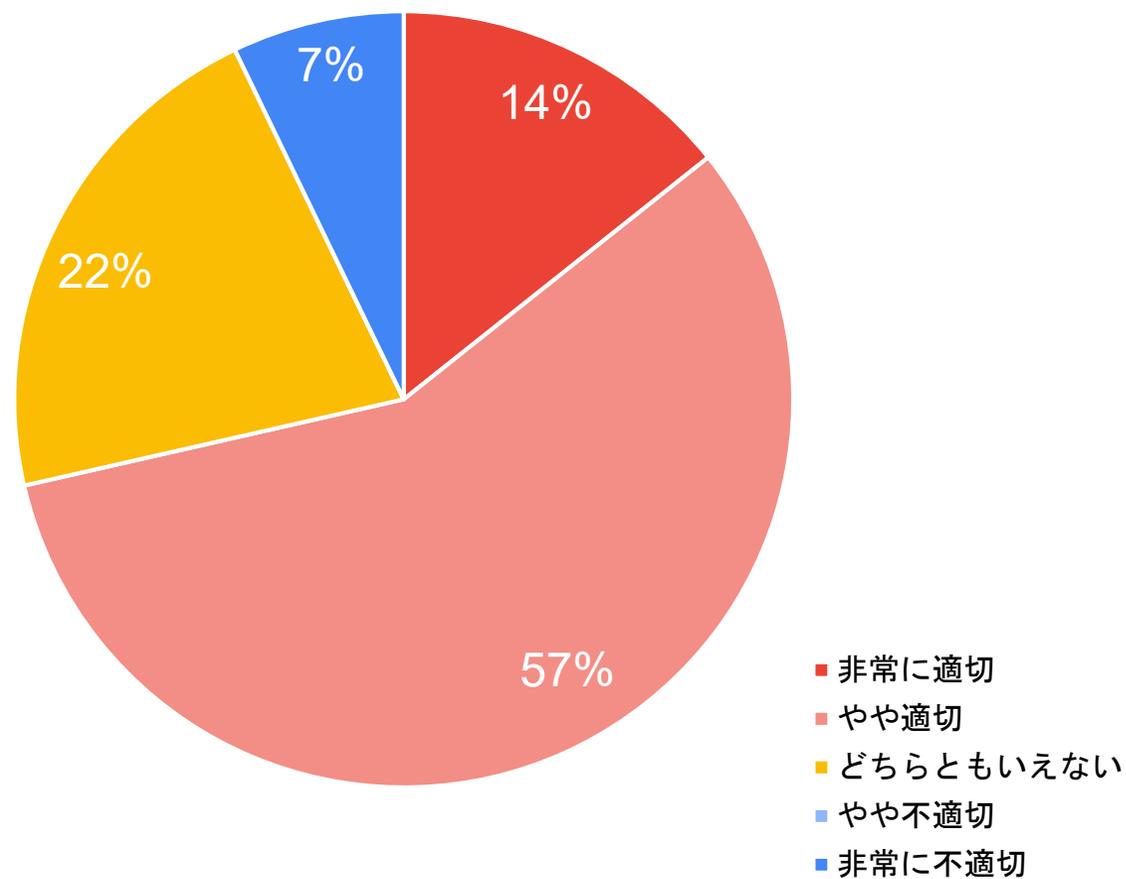
3. 理解のしやすさ



4. 自然さ



5. 倫理的配慮について



6. ご意見・ご感想（1部抜粋）

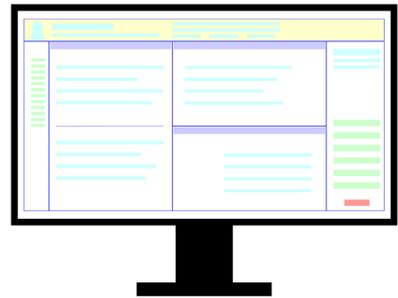
回答者	内容
1	<p>幽門側胃切除後の食道癌に対して残胃を温存する術式は、残胃に血流障害が生じる可能性が高く、臨床的には不可能ではないかと思えます。そのため、プロトコルの文面というよりも医学的な観点からAIが作成したプロトコルに対する評価が少し厳しめになった可能性があります。</p>
2	<p>文章がですます調とである調が混在していることが気になりました。また、例えば、15節の「・・・二次利用（メタアナリシスなど）が可能である。」など、プロトコルに記載する内容ではなさそうな文言があったり、逆に、Visit等のスケジュールがわかりにくかったりしているように思えます。一方で、CQに関する内容は読み取れるなど、校閲前のプロトコルという位置づけでは、これをたたき台にすることは不可能ではないようにも思いました。</p>
3	<p>生成AIは現時点では今回のように「膨らませる」方向よりも「まとめる」方向に活用するのが現実的。プロトコルからIC文書を作成したり、学会スライドからlay summaryを作成したり、と比較的リッチな情報源から、別の形式に変換する方が向いているのではないかと。例えば背景に記載すべき本試験のrationaleは極めて弱く、インターネットから適切にデータを切り取って、inputした仮説に対する背景を充実させるというところは明らかに発展の余地がある。このあたりはプロンプトで改善できる余地があるのかどうか興味深い。</p> <p>また研究代表者や研究組織など、明らかなハルシネーションと思われる部分も散見される。プロトコルでの活用を考えると、改変の必要のない章はテンプレート通りに記載するようプロンプトで指定してもよいかもしれない。</p>

テーマA：最終的な評価

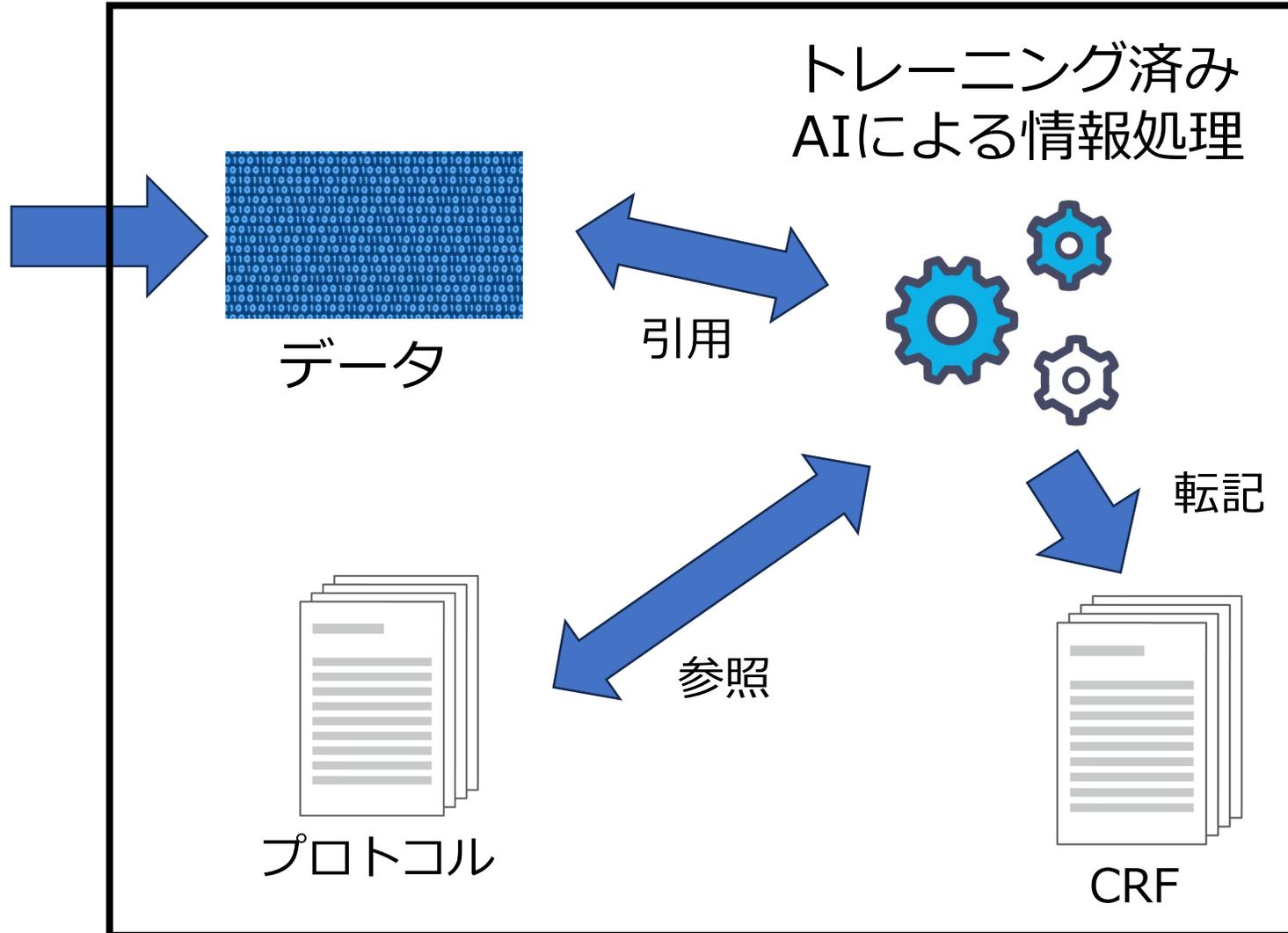
- 改善を試みることにより、かなり改善されたが、本来、考えてほしい Clinical Questionまで到達しなかった
- 生成AI（DIP社 WritingX）そのものが手術プロトコルを書いたことがなかったなので、難しかった＊医薬品プロトコルを書かせたところ、5万字、46ページのプロトコルを書き上げた
- 今回、シノプシスを基にプロトコルを書かせたところ、文体や表現の一貫性、内容の整理と明確化、倫理的配慮と同意手続きなどで、改善すべき点が存在することが明らかになった。今後これらの改善により生成AIによるプロトコル作成がより高度なレベルで実用化できる可能性が示唆された
- AIが生成するプロトコルは、研究者にとって有用な「たたき台」や「参考資料」となり得るため、有用な支援ツールとなり得ると考えられた
- 今後は、章ごとに分割してライティングをすることにより、特にお作法部分に関しては、かなりの改善がみられるのではないかと思われた

テーマC 電子カルテ情報からの自動情報抽出精度の検証

生成AIソフトウェア（DIP社）



電子カルテ
情報



転記の精度に関する
検証を行う

方法

- 過去の治験プロトコルを参照して、選択除外基準部分のCRFを作成する
- 電子カルテのデータ（、画像診断報告書、患者プロフィール、患者基本/患者属性、看護サマリ、電子カルテの記載、上部内視鏡検査報告書、熱型表バイタル記録リスト、臨床検査値一覧を抽出対象）を引用して、マッチングさせてみて、Yes/Noを記載させる
- 対象としては、過去に消化器外科領域で実施した医師主導治験（15例）とした
- 生成AIは電子カルテデータから、Yes/Noの判別理由と共にYes/Noを記載させる

結果

- 適格基準の判定精度に関して、15名のうち14名(93.3%)で6項目全てを正しく判定した。1名のみ、「他に活動性の重複癌（血液悪性腫瘍を含む）がない患者」の1項目で基準を満たさないと判定された（全体精度98.9%）
- 除外基準に関しては、対象患者の判定項目合計数120のうち、除外基準を満たすと判定された項目が1（0.8%）、除外基準を満たさないと判定された項目が107（89.2%）、判定不可の項目が12（10.0%）であった

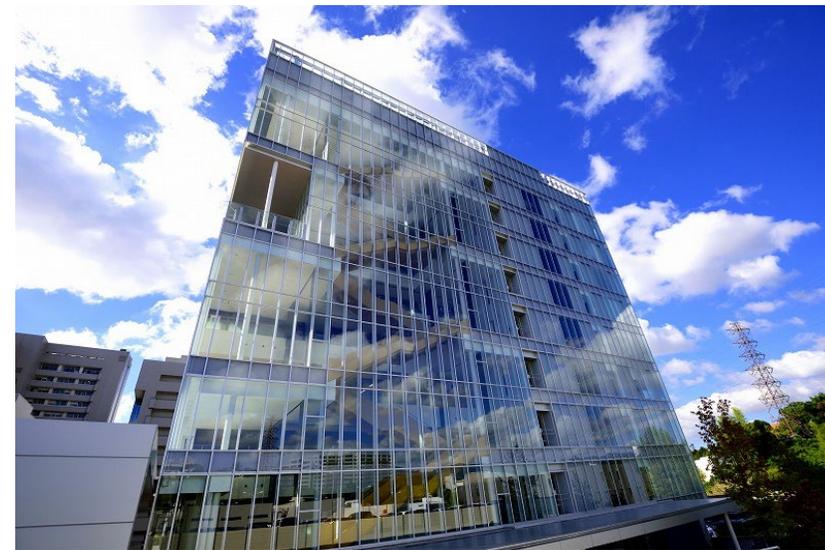
テーマC：最終的な評価

- 生成AIを使用して電子カルテデータからeCRFへの自動入力や、治験実施計画書からのeCRF項目の抽出が可能であることが示された
- 適格基準で誤った判断を下した1例は、根拠として「リンパ腫の合併」が挙げられていたが、実際にはそのような併存症はなかった。腹部造影CTレポートに記載された「リンパ節腫大」や「リンパ節転移」が誤判定の原因だと推察される
- 除外基準に該当すると判定された一例では、「コントロール不良な感染症（活動性の結核を含む）を有する患者」が原因であった。具体的には術前の臨床検査でHBs抗原陽性、HBs抗体陰性、HBc抗体陽性と判定された活動性B型肝炎の患者であった。「コントロール不良な」という表現の曖昧さが誤判定を引き起こした可能性がある
- 「妊娠、授乳婦である患者」に関しては、性別から明確に除外基準を満たさないと判定できる場合もあるが、本検証ではカルテの構造化データや記録には根拠が存在しなかったため、「判定不可」と出力されることが望ましいが、そうではないものもあった
- 精度と信頼性を向上させるためには、生成AIによる判断根拠を明確に提示しそれを容易に確認できるユーザーインターフェースの設計が重要である

まとめ（私論）

- 生成AIによる治験/臨床研究関連文書に関しては、提出文書と言えるまでのレベルではないが、十分に検討資料として扱えるレベルであることがわかった
- 電子カルテからのデータ抽出という観点では、精度高く実施することが出来たが、AIとの共存という意味では、人とAIが理解していることの共有が必要であると思われた
- 今後は、「人」と「AI」がどう共存していくかを考える必要があり「人」対「AI」ではなく、「人」対「AIを使う人」となるのではないかと思われる

For the Patients and All Stakeholder's good life



By BingAI (DALL-E 3)

ご清聴ありがとうございました

- ご質問、お問い合わせは以下まで
- e-mail : k-asano@dmi.med.osaka-u.ac.jp



本発表はAMEDの課題番号JP17lk1503002による成果も含みま