

厚生労働科学研究費補助金（厚生労働科学特別研究事業）
分担研究報告書

テーマC：電子カルテからの自動情報抽出精度の検証

研究代表者	浅野健人	国立大学法人大阪大学	医学部附属病院未来医療開発部・特任准教授
研究分担者	武田理宏	国立大学法人大阪大学	医学部附属病院医療情報学・教授
研究協力者	和田聖哉	国立大学法人大阪大学	医学部附属病院医療情報学・寄附講座助教
研究協力者	山本晋也	国立大学法人大阪大学	医学部附属病院未来医療開発部・招へい教授

研究要旨

目的：本研究では、自然言語およびテンプレート形式、検査記録などを含む電子カルテのデータを用いて、生成 AI により模擬研究 electric CRF (eCRF) の入力項目に自動的にデータを抽出することができるかどうかとその正確性の検証、また自動抽出された eCRF の入力内容から模擬研究の参加適格者/不適格者の仕分けを適切に実行できるかどうかを検証することを目的とする。

方法：2015年1月～2023年12月に大阪大学医学部附属病院に通院されていた患者の電子カルテデータを情報源として、DEEP INTELLIGENT PHARMA(SG) PRIVATE LIMITED が提供している WritingX 等の生成 AI ソフトウェアを用いて、模擬研究 eCRF の自動入力の実験と参加適格者/不適格者判定の実験を実施した。

結果：模擬研究 eCRF の自動入力の実験では 99%を超える精度を示したが、記載のない情報に関するハルシネーションが発生することを確認した。模擬研究で参加適格者/不適格者の判定を行う実験では、適格基準の判定精度は 98.9%を達成したが、除外基準では特定不可と除外対象外の判定が混在した。

結論：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 項目の抽出する可能性を示すことができた。判定結果の信頼性を向上させるためには、その判断根拠もアプリケーション上で明確に提示することが重要となる。

A. 研究目的

臨床研究を行う際に、研究に関する情報を研究用のデータとして収集する際には医師、臨床研究コーディネーターが電子カルテを閲覧しながら electric CRF (eCRF) に必要な情報を入力する作業を担っている。研究を担当する医療従事者はその情報入力作業の多くを手作業で行う必要がある。電子的に情報入力を行う electric data capture (EDC) が普及しているとはいえ、電子カルテに保管された情報をそのまま電子的に抽出、引用することは困難であり、入力作業は未だ手作業

で行われている。本研究では自然言語およびテンプレート形式、検査記録などを含む電子カルテのデータを用いて、生成 AI により模擬研究 eCRF の入力項目に自動的にデータを抽出することができるかどうかとその正確性の検証、また自動抽出された eCRF の入力内容から模擬研究の参加適格者/不適格者の仕分けを適切に実行できるかどうかを検証することを目的とする。

B. 研究方法

情報源としては、2015年1月～2023年12

月に大阪大学医学部附属病院に通院されていた患者の電子カルテデータを用いる。事前に個人情報や個人が特定できる情報、診療に関する機微な情報などが含まれていない形に限定した。

1. 模擬研究 eCRF の自動入力検証

生成 AI が電子カルテの診療録から指定した項目に従って正確に情報抽出を行うことが可能かどうかを検証するために、大阪大学医学部附属病院で採用されているダイナミックテンプレートを用いて記録された電子カルテ情報を利用した。本テンプレートから入力すると、その内容が構造化データとして格納される。また、診療記録上で表示される際には違和感のない自然な文（自然言語文）に変換して出力されるため、これらを組み合わせると自然言語文と構造化データのペアが準備できる。本検証では、生成 AI に自然言語文と構造化データの項目名を与え、構造化データの値が正しく抽出できるかを評価することで、eCRF の自動入力を擬似的に検証することとした。

対象患者は心不全の増悪で入院し、患者背景や既往歴などを含む下記 42 項目に欠損のない 50 名を対象とした。各項目の詳細とデータ型は以下の通りであった：

- 日付：3 項目
入院日、身長測定日、体重測定日
- 数値：10 項目
入院時年齢、身長、体重、Body Mass Index、来院時収縮期血圧、来院時拡張期血圧、心拍数、体温、経皮的動脈血酸素飽和度、心胸郭比
- 文字列：5 項目
性別、入院目的、入院形式、NYHA、胸部エ

ックス線体位

- カテゴリ（あり/なし/不明）：24 項目
急性心不全による入院、過去の心不全入院治療歴、植込みデバイス、在宅酸素療法、持続陽圧呼吸療法、自動調節サーボ換気、補助人工心臓、維持透析、経皮的冠動脈インターベンション/経皮的冠動脈形成術、カテーテルアブレーション、経皮的心臓手術、開心術、高血圧、糖尿病、脂質異常症、慢性閉塞性肺疾患、心房細動/心房粗動、慢性腎臓病、肝疾患、膠原病、喫煙歴、多量飲酒歴、入院時 6 分間歩行テスト、入院時 Fibroscan

本テンプレートではこれら以外にも現病歴やその他の既往歴、治療方針などが入力されており、それらも含めて自然言語文に出力される。本検証ではこれらの情報を誤って抽出対象としないかどうか評価に含まれる。また、上記 42 項目のうち、性別は自然言語文の出力には含まれない。本項目の出力を他の 41 項目とは区別して評価し、生成 AI が記述されていない情報を出力してしまうハルシネーションについて検討した。

本検証では生成 AI として東京リージョンの Amazon Bedrock で利用できる Claude 2.1 を採用した。プロンプトでは自然言語文と各項目名を与え、JSON 形式で解釈可能な文字列を生成する指示を構築した。なお、各項目の医学的説明は指示プロンプト内では与えていない。

2. 研究参加適格者/不適格者の仕分け検証

DEEP INTELLIGENT PHARMA(SG) PRIVATE LIMITED が提供している WritingX の生成 AI アプリケーションを用いて、治験実施計画書から模擬研究の参加適格基準/除外基準の自

動抽出を試みる。また、電子カルテデータ(診療録、検査報告書、検体検査結果、入院時バイタルサインなど)を書き下した自然言語文を生成 AI アプリケーションに与え、自動抽出された各基準項目について適格/不適格/判定不可の判定をその理由とともに出力し、適切に判定できているかを検証する。

・電子カルテデータ提供のプロセス

対象患者は大阪大学医学部附属病院で過去に実施された治験(腹膜播種を評価するための審査腹腔鏡検査に関する治験)から選択し、15名を選出した。模擬研究の条件は同治験の条件を参考にして以下の項目を設定した:

【適格基準】

- 1) 同意取得時の年齢が、20歳以上の患者。
- 2) 進行胃癌のうち大型3型(最大径8cm以上)、4型胃癌もしくはCT画像上で腹膜播種が疑われる胃癌患者。
- 3) ECOG Performance Statusが0~1の患者。
- 4) 心、肺、肝、腎、骨髄機能が保たれており、臨床検査値が以下の条件をすべて満たす患者。(a~c、eについては、審査腹腔鏡施行前14日以内の検査値を用いる。dについては、同意取得前14日以内の検査結果の利用も可能とする。) a. 腎機能: BUN \leq 施設正常値上限の1.5倍 血清クレアチニン値 \leq 施設正常値上限の1.5倍 b. 肝機能: AST(GOT) \leq 施設正常値上限の1.5倍 ALT(GPT) \leq 施設正常値上限の1.5倍 血清総ビリルビン値 \leq 施設正常値上限の1.5倍 c. 血液: 好中球 1,500/mm³ 以

上血小板 100,000/mm³ 以上 Hb 8.0g/dL

以上 d. 心機能心電図: 重篤な異常を認めない e. 肺機能: 肺機能: 動脈血酸素飽和度(SpO₂) \geq 94% (酸素非投与下)

- 5) 他に活動性の重複癌(血液悪性腫瘍を含む)がない患者。
- 6) 本人に病名・病態の告知がなされており、かつ本人に同意能力があり、文書同意が得られた患者。

【除外基準】

- 1) 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。
- 2) 重篤な合併症(悪性高血圧、重症のうつ血性心不全、重症の冠不全、同意取得前3ヶ月以内の心筋梗塞、末期肝硬変、重症の腎不全あるいは血液透析を要する慢性腎不全、コントロール不良な糖尿病、重症の肺線維症、活動性の間質性肺炎等)を有する患者。
- 3) 同種骨髄移植後の患者。
- 4) 重症の精神障害を有する患者。
- 5) ポルフィリン類縁物質に対して過敏症又は既往のある患者。
- 6) 他の治験(臨床研究を含む)に参加している患者。
- 7) 妊娠、授乳婦である患者。
- 8) その他、治験責任医師又は治験分担医師が不適格と判断した患者。

これらの情報が記載されている可能性のある電子カルテデータとして、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とした。一定のフォ

ーマットで自然言語文に変換できるようにするために、各情報源をシート名にして、構造をできる限り保持したxlsx ファイルを患者単位で作成した。xlsx ファイルを研究協力者へ転送する前に、同ファイルをアジャイル暗号化して提供した。

- ・生成 AI アプリケーションの詳細説明

プロトコル解析:プロトコルの文書情報を構造化解析し、選択基準/除外基準、試験日、名称、基本情報などを抽出した。

データ前処理:被験者データに対してプライバシー保護処理を実施し、個人を特定できる情報をチェック、マスキングすることでデータの安全性を確保した。また、Day 0 以前の情報のみを参照するようにデータをフィルタリングした。

自動判定処理:大規模言語モデル (LLM: GPT-4 turbo) とカスタムプロンプトを用いて、被験者データとプロトコル情報を分析し、被験者が選択基準/除外基準を満たすか否かを自動的に判定した。また、ユーザーが判定結果を最適化するために、プロンプトや LLM 関連パラメータ (モデル選択、温度設定、最大応答長など) をカスタマイズできるように UI を設計した。そして、被験者が選択基準/除外基準を満たすか否かを出力し、その判定理由も併せて提供した。また、結果の出力を可能にすることで、その後の分析やアプリケーションに活用できるように機能を追加した。

- ・倫理面への配慮

本研究は、電子カルテ情報を用いて実施する観点から、本研究実施計画書を作成し、大阪大学医学部附属病院倫理審査委員会によ

る審査を受け実施した (承認番号: 23477)。

C. 研究結果

1. 模擬研究 eCRF の自動入力検証

50 名の自然言語文の平均文字数は 2,941 ±802 であった。性別を除いた 41 項目の精度は 99.7% (2,044/2,050) を達成した。一方で、性別に関しては「記載なし」もしくは空白が出力されたのは 12% (6/50) にとどまった。また、7% (7/50) で逆の性別が出力された。

2. 研究参加適格者/不適格者の仕分け検証

治験実施計画書は A4 全 38 ページで構成され、そのうちの 2 ページにわたって適格基準と除外基準が記載されていた。生成 AI アプリケーションは治験実施計画書から適格基準 6 項目、除外基準 8 項目の全てを正しく自動抽出できた。

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。

D. 考察

1. 模擬研究 eCRF の自動入力検証

自然言語文において項目名や値が明確に記載されている場合、自動構造化は極めて高い精度 (99.7%) で実行可能であることを確

認できた。また、今回の検証で欠落した6項目は身長もしくは体重の測定日であり、いずれも生成AIの出力は身長・体重の測定値の後方に測定日が記述された形式となっており、完全な見落としではなかった。一般的に生成AIでは、学習に用いたテキストのドメインやタイプと異なるタスクに対してはパフォーマンスが低下する可能性が懸念されている。しかしながら、本検証結果は、専門的な医学用語が含まれる日本語のプロンプトでも生成AIが有効に機能して情報抽出を達成したことを示している。これは、臨床データの自動抽出において重要な進歩を示唆する。

一方で、性別項目におけるハルシネーション現象が観察され、生成AIが文脈から性別を推測する際に不正確な情報を生成しうることが明らかになった。この問題に対処するためには、生成AIの推論根拠を明確にするとともに、ユーザーが生成AIの判断を容易に評価できるような仕組みの導入が重要である。

本検証の限界は、入力が自然言語文とはいえ、特定のフォーマットに沿って記述された文章であり、再構造化が比較的容易である点にある。より自由度の高い文章に対して、各判定項目に関して医学的解釈を要する場合の検討は今後の課題である。

2. 研究参加適格者/不適格者の仕分け検証

生成AIアプリケーションは治験実施計画書から適格基準および除外基準を正確に抽出し、またそれらの項目毎に判断根拠を提示しながら高精度に判定を出力することができた。しかし、適格基準で誤って不適格判定された患者が一例存在した。この患者の判断

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。

一方、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。

なお、除外基準に関して、今回の対象者は実際の治験に全員参加できているため、臨床的には除外基準を満たしていないという判定は正しいと言える。しかし、これらの項目を判定するための具体的な情報は電子カルテの構造化データには存在せず、しばしば電子カルテの記録にも記載されないこともある。「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、本検証ではそれ以外の除外基準については「判定不可」と出力されることが望ましい。除外基準を満たさないと判断された項目の判断根拠を確認すると、「no mention」や「indicated」のような表現が使用されていた。生成AIアプリケーションは直接的な記載や情報がない場合の解釈も行っているようにみえるが、それを「特定不可」と明確にするのは困難な問題だと考えられる。

本検証の限界として、模擬研究で参照した治験に参加した患者以外での評価が行えていないことが挙げられる。適格基準を明らかに満たさない患者や、除外基準の疾患を合併

している患者に対して生成 AI アプリケーションが下した各項目の判定と判断根拠を精査することで、記載のない情報に対して高精度に「特定不可」と出力させることのできる指示プロンプトが構築できる可能性が残されている。

E. 結論

本研究により、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。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。

F. 健康危険情報

なし

G. 研究発表

1. 論文発表
なし
2. 学会発表
山本 晋也 「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」日本臨床試験学会第 15 回学術集会総会 2024/3/8

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