

がん診療支援のための大規模言語モデルの研究開発

1. 研究の概要と目的

研究は、国立がん研究センター研究倫理審査委員会および国立がん研究センターの長の許可を受けて実施するものです。

近年、大規模言語モデル(LLM)と呼ばれる人工知能技術が急速に進歩しており、医療分野での活用が期待されています。特に、複雑で膨大な情報を取り扱うがん診療において、LLMは医師の意思決定を支援する有力なツールとなる可能性があります。しかし、一般的なLLMは専門的な医学知識が不十分であったり、事実に基づかない情報を生成する「ハルシネーション」という課題を抱えており、そのまま臨床現場で利用することはできません。

そこで本研究では、日本の医療現場で安全かつ有効に利用できる、がん診療に特化した大規模言語モデル(LLM)を開発することを目指します。

その一環として、過去に国立がん研究センター中央病院でがんと診断され、診療を受けられた患者さんの電子カルテ情報を活用し、開発したLLMの性能を評価するための高品質な「質問応答データセット」を作成します。このデータセットは、実際の臨床現場で想定される質問と、それに対する専門医の模範的な回答で構成されます。

本研究を通じて、医師の負担を軽減し、将来の患者さんにより質の高いがん医療を提供するための基盤技術を創出します。

2. 研究の対象となる方

2000年01月01日～2027年03月31日までに、国立がん研究センター中央病院において、がんと診断され診療を受けた20歳以上の患者さん。

上記に該当する方のうち、最大500名分の診療情報を参照させていただきます。

3. 研究目的・方法

研究実施期間:研究許可日から2027年3月31日まで。

本研究では、まず学術論文や診療ガイドラインなど公開されている情報を学習させ、がん診療に特化したLLMを開発します。

次に、研究の対象となる患者さんの診療情報を参照し、臨床現場を想定した実践的な質問と専門医による回答からなる「質問応答データセット」を手動で作成します。

患者さんの個人情報を含む元の診療データは、LLMの学習には一切使用せず、開発したLLMの性能評価(精度や安全性の検証)にのみ利用します。

4. 研究に用いる試料・情報の種類

本研究で利用させていただくのは、電子カルテに記録された以下の診療情報です。生体試料(血液や組織など)は一切利用しません。

診療テキスト情報: 患者さんの性別、年齢、疾患名、ステージ、治療内容、経過記録、病理レポート、放射線レポート、退院サマリー、化学療法レジメン情報、予後等の情報

5. 個人情報の取り扱いについて

本研究では、患者さんのプライバシー保護を最優先し、個人情報の取り扱いに細心の注意を払います。診療情報を参照してデータセットを作成する際には、以下の手順を徹底します。

匿名化の徹底: 氏名、診療録 ID、生年月日、住所など、個人を直接特定できる情報はすべて完全に削除します。

情報の加工: 検査値や薬剤の投与量などは、臨床的な意味を損なわない範囲で、値を丸めたり、意図的にノイズを加えたりする加工を施し、個人が特定されるリスクを最小化します。

対応表の不作成: 匿名化したデータと、元の患者さんの情報を結びつける対応表は作成しません。これにより、万が一の場合でも個人を再特定することは不可能です。

希少例の除外: 特定の個人に結びつく可能性のある、極めて稀な症例に関する記述はデータセットに含めません。

作成した質問応答データセットの全部または一部を、学術論文などで公開する可能性があります。その場合でも個人が特定できる情報は一切含まれません。

6. 研究組織

研究責任者

国立がん研究センター研究所 医療 AI 研究開発分野 主任研究員 小林 和馬

研究分担者

国立がん研究センター研究所 医療 AI 研究開発分野 特任研究員 柴木 亮太

国立がん研究センター研究所 医療 AI 研究開発分野 外来研究員 松本 大海

国立がん研究センター中央病院 頭頸部外科 医員 坂井 梓

7. お問い合わせ先

本研究に関するご質問等がありましたら、下記の連絡先までお問い合わせください。ご希望があれば、他の研究対象者の個人情報および知的財産の保護に支障がない範囲内で、研究計画書および関連資料を閲覧することができます。なお、この研究が適切に行われているかど

うかを確認するためや研究の科学的意義を検討するために、国内外の第三者の立場の者が研究対象者の方のカルテやその他の診療記録、研究データなどを拝見することがあります。このような場合でも、これらの関係者には守秘義務があり、個人情報は守られます。

また、ご自身の診療情報が本研究に利用されることを希望されない場合は、下記までお申し出ください。お申し出いただいた場合、あなたの情報を研究に利用することはありません。なお、利用を拒否された場合でも、今後の診療等で不利益を受けることは一切ありません。

【照会先および研究への利用を拒否する場合の連絡先】

〒104-0045 東京都中央区築地 5-1-1

国立研究開発法人 国立がん研究センター研究所 医療 AI 研究開発分野

担当者:小林 和馬

電話番号:03-3542-2511 (内線: 3663)