[redacted]

月経血幹細胞と脂肪幹細胞研究におけるAIを使用した文献解析の精度と進歩

伊沢 博美 1 、小川 誠司 1,2 、吉居 絵理 1 、小谷 佳織 1 、永田 明久 1 、森田 祐二 3,4 、久永 一郎 5,6

1[redacted]2藤田医科大学東京先端医療研究センター

3ホルムクリニック 、4同志社大学生命医科学部アンチエイジングリサーチセンター/糖化ストレス研究センター

5立命館大学アート・リサーチセンター 、6大日本印刷株式会社マーケティング本部ソーシャルイノベーション研究所共創ビジネスデザイン部ヒューマン・エンジニアリングラボ

Abstract

【目的】 近年、ChatGPTは科学文献検索における有用性が高く評価されている。本研究では、特に再生医療分野における文献検索においてChatGPTの精度と進歩を検証し、月経血幹細胞(MenSC)と脂肪幹細胞研究(ADSC)への応用の可能性を探る。

【方式】 本研究では、神宮外苑Wom anLifeClinicでインフォームド・コンセントを取得し提供している再生医療を中心として、ChatGPT-3.5とChatGPT-4を使用した文献検索の精度を比較検討した。MenSCは、体外に排出されることから、月経カップを用いた非侵襲的採取が可能である。また、MenSCの静脈投与および局所投与の安全性についても、奇形腫・異形成・免疫反応を引き起こすといった報告も見られず、優れた幹細胞源の1つと考えられているが、2007年からと文献が少ない。一方、ADSCは、その豊富な存在量や分化能からも、組織修復において重要な役割を果たすことが知られており、国内でも広く治療が提供されており、文献も豊富にある。今回我々は、MenSCおよびADSCに関連する専門的な文献データベースを用いて、ChatGPTによる検索結果の正確性と有用性を評価した。

【結果】 初期のChatGPTモデルによる文献検索結果には誤った情報の問題が存在したが、ChatGPT-4の登場により、その精度は大幅に向上している。

【結論】 本研究により、ChatGPTは特定の科学的分野における文献検索において有用なツールである可能性が示された。しかし、その精度は使用するバージョンや設定、さらには使用者の具体的な指示に大きく依存する。したがって、ChatGPTによる検索結果は、追加で検証を行うことが重要である。将来的には、このツールのさらなる改良により、再生医療研究における文献解析の精度と効率が飛躍的に向上することが期待される。

Introduction

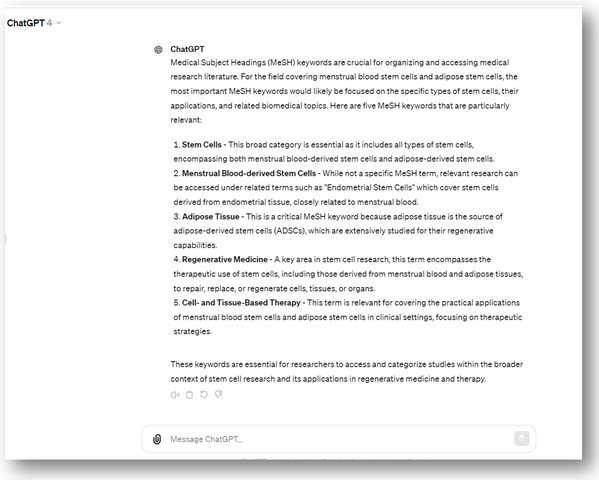
月経血幹細胞による治療は、非侵襲的に採取可能で、世界的にもさまざまな疾患への応用が期待されていることが知られている。一方、更年期症候群は、加齢に伴う変化の1つとして認識されており、女性ホルモンの大幅な変動とその後の低下が主要な原因とされており、加齢、心理的要素、ストレスなどが複合的に関与し、多様な症状を引き起こし、心身の健康感におけるQOL（Quality of Life）の低下をもたらす。神宮外苑Woman Life Clinicでは、更年期症候群を含む加齢に伴う身体の変化に対して、包括的な細胞治療アプローチを提供している。これには、卵巣機能低下への対策として自家月経血由来幹細胞の静脈内投与、フレイル及びプレフレイル状態の加齢性変化に対する自家脂肪由来幹細胞の投与、老化に伴う免疫力低下に対するNK細胞療法が含まれる。さらに、不妊症に対する治療として自家月経血由来幹細胞を卵巣および子宮腔内に投与している。

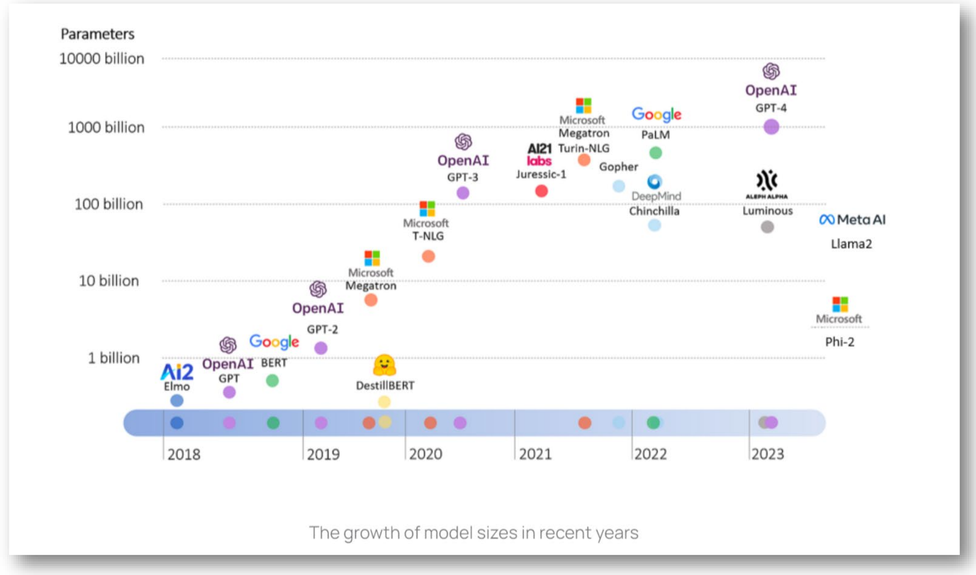
この研究の目的は、特に再生医療分野における文献検索においてChatGPTの精度と進歩を検証し、月経血幹細胞(MenSC)と脂肪幹細胞研究(ADSC)への応用の可能性を探ることである。

Materials and methods

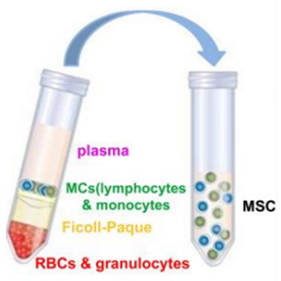
ChatGPT (Chat Generative Pre-Trained Transformer)　OpenAI　のGPT-3.5とGPT-4を用いて調査を行った。まず初めに、ChatGPT-4を用いて月経血幹細胞および脂肪幹細胞研究の分野で最も重要な5つのMeSH（Medical Subject Headings）キーワードを特定した。次に、GPT-3.5とGPT-4を使って「[MeSHキーワード]に関するエッセイを参考文献とともに書く」というプロンプトエンジニアリングでエッセイを作成した。それぞれのエッセイで生成された参考文献を収集し、重複を取り除いた後、手作業で分析した。5つのエッセイからGPT-3.5による17件とGPT-4による20件の合計37の参考文献が抽出された。

MeSH keyword ：Stem Cells , Menstrual Blood-derived Stem, Adipose Tissue, Regenerative, Cell- and Tissue-Based Therapy



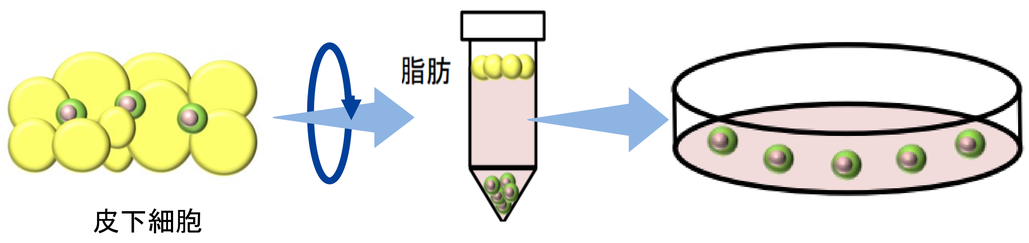


月経血幹細胞



- シリコン製のカップ
- 月経血採取は経血量の多い2日目に行う

脂肪幹細胞

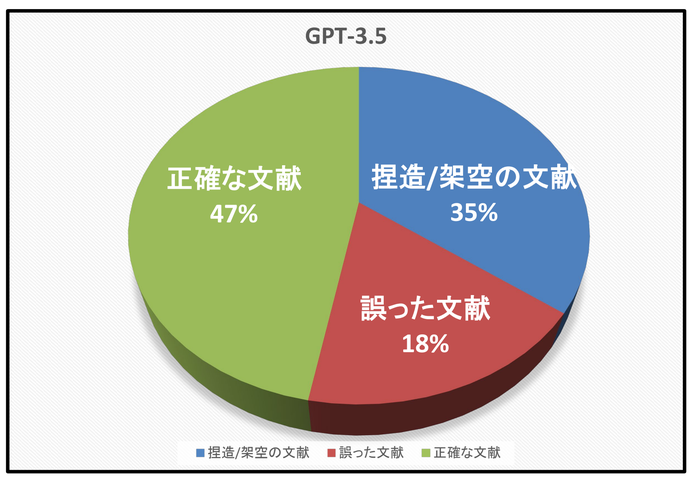


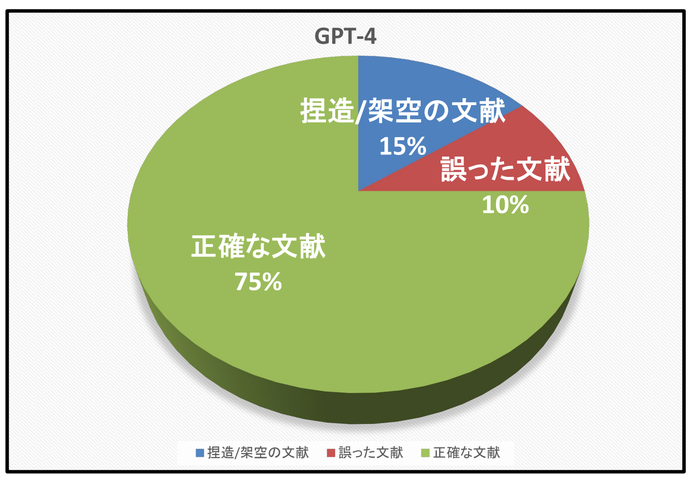
Result

抽出された文献のうちGPT-3.5による抽出17件とGPT-4による抽出20件のうち、GPT-3.5を用いた6件が捏造・架空の文献であり（35.3%、6/17）、3件は著者および出版年に誤りがあった（17.6%、3/17）。一方、GPT-4 を用いた3件が捏造・架空の文献であり（15.0%、3/20）、2件は著者および出版年に誤りがあった（10.0%、2/20）。

一方、2023年Wagner MWらの報告では、臨床放射線学的情報の検索にGPT-3を使用した情報と参考文献の精度は、生成された文献のうち36.2%のみがインターネット検索で入手可能な正確なものであり、残りの63.8%はChatGPTによって捏造されたものであったと報告されている。

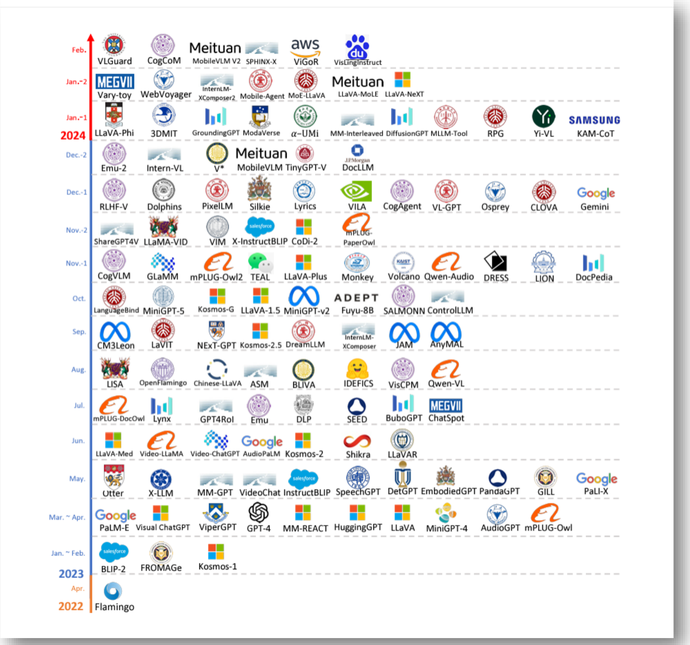
これらを鑑みると、捏造文献の発生率が高いのは、言語モデルのより古いバージョン（GPT-3>GPT3.5>GPT4）を使用したことと、データ収集に使用したトピックが原因であると考えられる。





Discussion

GPT-3.5とGPT-4は、従来のGPT-3に比べて捏造や架空文献の発生率が低い傾向が見られるが、新しいモデルでも完全に排除されているわけではない。特に、インターネット上の情報量に左右される点が重要です。ChatGPTはインターネット上に誤情報の方が多ければそれを反映してしまい、新薬が登場し治療の長い歴史が塗り替えられたとき、正しい情報に更新されるまでには時間がかかる。したがって、ChatGPTの生成する情報には誤りや不正確な部分が存在する可能性があり、それらを鵜呑みにするのは危険である。これを前提として医学分野でChatGPTを使用する際の課題を考察する。医師の診察行為は、患者の肌の色つや、表情、動作、主訴、声色、語気など、多くの情報を組み合わせるマルチモーダルな情報処理が求められる。しかし、現状のGPT-4はテキストベースで一度テキスト化された情報に限られる。画像や音声のマルチモーダル処理へのバージョンアップが待たれる。さらに、医学専門知識のファインチューニングについても、2022年を転機にChatGPTのように大量のリソースをつぎ込み大量のパラメータを高速に処理する拡大路線とは異なり、限られたパラメータの学習を行う小規模なMeta社やマイクロソフト社のAIが注目されている。これらの動きを勘案すると、現時点では何かひとつのAIだけを信頼するのではなく、複数のAIを同時に使用することで医療のニーズにハックしていくしかない。最後に、この報告に日本のAIをテストできていないことを大変残念に思う。大切な医療情報をできれば日本国内で留保できる仕組みを医療業界からも訴えていく必要を感じる。



References

1) https://www.predli.com/post/llm-deep-dive-phi-2
2) Can Assoc Radiol J. 2024 ;75(1):69-73.
3) arXiv preprint arXiv:2401.13601 (2024).
4) 月経血幹細胞静注投与：細胞数による治療効果の差. 第23回再生医療学会総会.ポスター(2024)
5) Cureus. 2023 ;15(7):e42150.
6) Clin Imaging. 2023 ;103:109993.

[redacted]

筆頭演者：　伊沢　博美
演題発表に関連し、開示すべきCOI関係にある企業などはありません。