

厚生労働科学研究費補助金(長寿科学政策研究事業)
「機械学習を用いた介護認定審査会の審査判定プロセス等を補助するシステムを開発するための研究」
分 担 研 究 報 告 書

医療分野における人工知能(AI)の現状に関する文献研究

研究分担者 久米 慶太郎 (筑波大学 医学医療系)
研究分担者 蜂屋 孝太郎 (帝京平成大学 人文社会学部)
研究分担者 福地 一斗 (筑波大学 システム情報系)
研究分担者 小宮山 潤 (筑波大学 医学医療系)
研究代表者 高橋 秀人 (帝京平成大学 薬学部)

研究要旨

【背景・目的】

本分担研究では、要介護認定審査における審査プロセスを補助する高精度な AI システムの構築に有用な情報を探るための文献調査を目的とする。

【方法】

直近 5 年の医療 AI の文献を調査し、技術的動向と本研究課題への適用可能性を検討した。

【結果】

医療 AI はルールベース、統計的学習を経て深層学習へと発展してきた。現在はハルシネーション、バイアス、知識の限界等の課題に対し、FT、RAG 等の手法で対策がはかられている。

【考察】

判定プロセスは現行の AI と親和性が高いと考えられ、RAG による根拠提示やルールベースと LLM のハイブリッド構成により、精度と信頼性の一定の担保が期待できると考えられる。

【結論】

本システムの開発においては、説明可能性と精度を両立し、審査会委員に受容されるようなシステムを構築することが重要である。

A. 研究目的

我が国は、急速な少子高齢化の進行により平成 19 年 1 月より超高齢社会（65 歳以上人口の総人口に対する割合が概ね 21%超）[1] に突入している。要介護（要支援）認定者数もその当時の 438.8 万人[2]から令和 7 年 12 月時点で 736.0 万人[3]に増加しており、この傾向は継続すると予想される。また、要介護認定の前段階にあたる、要介護認定に係る申請件数自体も 2022 年度における総数は 675.3 万件[4]となっており、認定審査に関与する関係者の事務的作業負担もまた増大して

いることを意味している。加えて、介護保険法第 27 条第 11 項において、市町村は被保険者から要介護認定を受けるための申請があった日から 30 日以内に処理しなければならないとされているが、令和 6 年度における要介護認定に係る事務に要する日数の平均値は 39.9 日（うち、認定調査所要期間 11.1 日、主治医意見書所要期間 17.9 日、介護認定審査会等事務処理期間 16.8 日、また、30 日以内に処理された件数の割合は 22.8%）[5]となっている。

このように要介護認定審査では、申請件数の増加や認定審査会委員の負担増大を背景とした自治体負担の軽減、審査の効率化、公平性の担保および地域間の均霑化の確保が喫緊の政策課題となっている。

このような経緯から、近年、介護認定調査における認定調査票作成支援、介護認定審査会に用いられる審査資料の読取・データ化、個人情報のマスキングなど、主に事務作業工程を対象としたAI活用が進んでいる。一方、要介護認定の多くの時間を要している二次判定および認定審査会における判定そのものを支援する仕組みは、認定調査票の特記事項や主治医意見書の自由記述を参照する必要のあるタスクであることによる困難さからか、いまだ実用化されていない。しかし、日数の短縮をはかるための介護認定審査会委員の増員も、労働人口の縮小からさらなる人材確保は困難であると考えられるため、現行の体制で迅速化・効率化をはかるには、AI等による技術的な支援が必須となる情勢となっている。

本分担研究は、医療分野におけるAIの応用事例を調査することによって、本研究課題で目標とする、「現場での利用において一定の評価を得られるような、認定審査会委員による二次判定を補助する高精度な判定支援システムの構築」に活用可能かどうか考察することを目的とした。

B. 研究方法

本分担研究の目的は、判定支援システムの構築に向けた技術的妥当性を検討することにあるため、個別の論文の分析ではなく、医療AIの動向を概観し把握することに主眼を置いた。調査にあたっては、PubMedを用いて特に直近5年間の原著論文およびレビュー論文を第一段階における調査対象とし、特に医療ドメインにおける傾向を概観し、その後第二段階として各論の調査を行った。

これにより、より実用化や現場に近いデータで検証実験・評価が行われている技術要素を効率的に特定し、本研究への適用可能性や考慮すべき問題点を考察するための資料とした。

(倫理面への配慮)

本分担研究は広く公開されている文献を対象とした調査研究であり、ヒトや動物に対する配慮を必要とする倫理的な問題は含まれていない。

C. 研究結果

人工知能の医療分野への応用の歴史

・ルールベース

医療分野において最初期に広く認知された成果は、スタンフォード大学のグループによって開発された抗菌薬治療の助言を行うプログラムである[6]。これはルールベースのエキスパートシステムとして開発され、あらかじめ定義されたルールに基づき、ユーザーとの対話的なセッションを通じて助言とその根拠となる説明を出力するプログラムであり、後にMYCINシステムへと発展した[7]。

また、この1970年代初頭には、精神科における妄想的な患者のシミュレート研究[8]や、生理学的な質問についてのデータベース応答を自然言語システムで行う研究[9]など、多岐にわたるタスクを遂行するプログラムが既に開発されている[6]。

・統計的学習

1990年代にはカーネルトリック等を組み込んだサポートベクターマシン[10-12]が急速に広まり、医療分野ではマイクロアレイデータに基づくがん細胞の分類などに初期応用例がある[13]。また、マイクロアレイのようなテーブルデータに対しては、1980年代に開発された決定木構築アルゴリズム[14]が、2000年代にはアンサンブル学習であるランダムフォレスト[15]へと発展した。ランダムフォレストの初期応用例としては、創薬分野においてマイクロアレイデータを対象としたものがある[16]。

さらに、決定木ベースの手法としては、勾配ブースティング[17]を利用したXGBoost[18]が特に有名であり、テーブルデータのような構造化医

療データにおいては、現在も強力な手法として採用され続けている。

一方、医療テキスト解析に関しては、一般的にはメール等のスパムフィルターなどの応用として知られる、ナイーブベイズ分類器を用いて医療テキスト(退院サマリー)に対して ICD コード(当時は ICD-9) を割り当てる研究例がある[19]。

・深層学習

いわゆるニューラルネットワークの基礎となる単純パーセプトロン[20]は 1950 年代に開発された。広く普及するのは 2010 年頃の深層学習として認知されて以後となるが、ただし、技術的な蓄積は続けられており、それ以前にも多層パーセプトロンを用いて ECG データから急性心筋梗塞を検出する研究[21]など多数研究例が存在した。大量のデータを取り扱うことができるようになった 2010 年頃に深層学習のパフォーマンスの飛躍的な向上が示され、初期は特に CNN による画像認識能力が認知されたため、医療ドメインにおいては画像への応用、例えば乳がん組織画像の解析[22]などが進んだ。

2018 年頃の BERT 登場後は、BioBERT [23]が開発され、医療テキストマイニングへの応用が示された。

さらに 2022 年頃、ChatGPT (GPT-3.5) が公開され、これを契機に診療への応用を検証する論文が急増した[24, 25]。

臨床分野別の傾向としては、画像データが多く特に親和性が高かった眼科や放射線科は LLM を用いた検証も多く、内科、医療情報も多い分野として挙げられている。一方、より致命的なリスクとなりうる疾患を伴う循環器科での例は相対的に少ないとされている[26]。ルールベースの人工知能の時代からしばしば応用が見られる精神医学の分野[27]も引き続き応用例が見られ、看護分野[28, 29]も多く見られるなど多様な臨床分野に及んでいる。

ただし、現状の検証の大部分は ChatGPT などの汎用 LLM によるものであり、医療 LLM の検証例は少ないと指摘されている[26]。

また、次なるパラダイムとして、現実世界との相互作用能力を持たせた物理 AI (PAI, EAI) の重要性も主張されている[30, 31]。

ユースケース別による応用の分類

臨床分野別ではなくユースケース別に応用を分類すると概ね以下の通りになる[28, 32]。

- 診断支援・医療知識検索
- 医療画像データ解析
- 患者モニタリング・管理・診断
- 生命医学 NLP (カルテやジャーナルなどからの情報抽出など)
- 創薬
- 臨床試験解析
- 医療 QA
- 医学教育および翻訳
- 医療製品の安全性監視
- 新興的な応用 (外科ロボットやソーシャル分析など)
- 性能検証

提案されている課題

現在の深層学習ベースの AI の、医療分野における問題点は概ね以下の 5 点にまとめられ、それぞれへの対応が検討されている[32-34]。

a. 限定的な知識・出力情報の不正確さ・ハルシネーション

この課題に対しては、

- RAG の使用
- インターネット検索機能の有効化
- 最新のデータセットを使用したファインチューニング(FT)
- 出力に対して信頼度または信頼性スコアを提供する技術の開発

のような対応が提案されている。

しかし、トレーニングには膨大な計算コストがかかり、FT も多くの計算コストを要し必ずしも個別の環境実施できるかどうかは不明で、なおかつ、再トレーニング・調整まで知識の更新はなされないで根本的な解決とはいえない[32-34]。

また、オープンウェイトモデルと独自 FT モデルの性能差は急速に縮小しているため[34]、かならずしもコストに見合うとは限らない。

そのため外部知識に接続してそれを LLM の出力に利用する技術である RAG の検討の価値があるとされている[35, 36]。

ただし、RAG を用いたとしても無視できない程度の誤答が依然として含まれることが指摘されている[35]。

したがって、RAG や FT などの個別ケースへの対応よりも適応戦略を洗練させることや、比較研究をより良くするための評価指標の整備が重要と指摘されている[36]。

その他、モデルから出力される誤りであるがもっともらしい文をハルシネーションと呼んでいるが、むしろこれは「作話」であるとの指摘がある[37]。

b. バイアスと公平性

これには、

- 多様なトレーニングデータセットを使用する
 - 多様なベンチマークを作成し、適用する
- のような対応が提案されている。

例えば、薬剤の処方予測について性別によるバイアスが含まれており[27]、意図的に設計しなければ公平なモデルは達成されないとしばしば指摘されている[27]。

また、RAG ではこのバイアスは解消されないことが指摘されている[36]。

c. 説明可能性

これには、

- 画像ベースのモデルに対しては可視化ツールを開発

- テキストベースのモデルに対しては内部概念を説明できる手法を開発

のような対応が提案されている。

また、判断根拠を明示できる RAG は説明可能性の向上に寄与すると期待されている[36]。

d. データセキュリティとプライバシー

これには、

- 匿名化フレームワークの構築
 - 研究データ管理計画 (DMP) を通じてデータセキュリティを保護
- のような対応が提案されている。

RAG や FT が LLM の知識の限界への対処法として提案されているが、個別環境で実施される場合、個人情報を含むデータセットに対するプライバシー保護の監視が行き届かない恐れがあり、データの漏洩のリスクが高まる懸念がある[36]。

e. 倫理・盗用・著作権・責任の所在

これには、

- リスクベースの規制フレームワークの適用
 - 国家機関、医療機関等によるモデルの継続的な監視・ガイドラインの提供
- のような対応が提案されている。

また、AI の出力の不正確さやバイアスといった技術的な欠陥や問題に対して、医療者側がそれを軽減するよう AI を管理・スーパーバイズすることが当然であるかのように技術者側が考えたり期待したりするのは極めて問題であるとの提起がなされている[37]。

D. 考察

これまでの医療分野における人工知能の応用の歴史と現状の課題から得られた示唆をもとに、要介護認定の二次判定支援システムの構築に向けて留意すべき点は何か、どう適用しうるかの可能性について考察する。

非構造テキストデータを入力とする医療 AI は近年急速に実証段階に入った分野である。要介護

認定の審査プロセスにおいて、特記事項や主治医意見書などのテキストデータの解析を必要とする本研究課題が対象とするタスクは、現在の AI 技術が特に貢献しうる親和性の高いものだといえる。

また、AI による支援を導入する際には、AI の判断の妥当性を評価するという新たなタスクの発生によってかえって負担が増加するというリスクがある。しかし、二次判定審査プロセスは一次判定の妥当性を判断するという性質をもち、AI による支援を導入しても本来のタスクとの方向性が大きく変わらないため、審査会委員の新たな負担増を抑えられると期待できる点においても望ましいものであるといえる。

さらにシステムの実効性について、実際のユーザーとして想定している層に、実際に使われることの重要性が提言されており、導入のハードルの高さは長年の課題である。ChatGPT の登場が汎用 LLM を用いた検証研究の増加[24, 25]に寄与していると考えられる一方、医療特化 LLM の検証例は相対的に少ない[26]ことから、非情報系ユーザーでも気軽に試すことのできる環境が実効性の点で重要であることが示唆される。したがって、本研究課題でも、限定的な条件下においてのみ有用なシステムを構築するのではなく、実地で一定の評価を獲得できるシステムとして開発できるかどうか極めて重要なポイントになると考えられる。

実効性については、データセキュリティとプライバシー保護の要件にも関わってくる。取り扱いに注意を要する機密性の高いデータを扱う都合上、外部から遮断された隔離環境で、なおかつ小規模な計算資源での運用を想定している本研究課題のシステムでは重要なポイントである。多くの検証や開発は潤沢な計算資源・環境の下で行われている[6, 7, 36]。エキスパートシステムの時代からパフォーマンスで高い評価を得ていても、計算資源の問題で使われないことは課題になっているため、計算資源が限られているエッジ環境での検証は重要であると考えられる。

この RAG については、現行の LLM の懸念点の一つであるハルシネーションを低減し、説明可能性を向上させるだけでなく、判断根拠を明示させることで、審査の公平性の担保に寄与すると考えられる。これは、初期の人工知能開発において指摘されるところの、専門家に効果的に利用してもらうためにはシステムが専門家と同じ方法で意思決定に至っていると確信させることが重要[6]というアプローチとも共通するものである。

また、実際の審査会委員による本システムの評価実験を通じてフィードバックを収集することで、LLM に由来するバイアスを分析し、より公平なものに改善できる可能性がある。さらに、このシステムを使う過程で、AI の出力と審査会委員の判断との相違を分析することで、LLM 由来のバイアスではなく、審査会間の相違が可視化できる可能性もある。

一方で、RAG 等の技術を用いてもバイアスや誤回答を完全に防止することは困難である。ユーザーが AI の出力を検証するとしても、審査の労力が単に AI の出力結果の検証という別の労力にかわるだけであってはならない。また、AI の出力に過度に依存する自動化バイアスやパフォーマンス低下の懸念も指摘されている[38]。

したがって、判断の責任の所在を明確にしつつ、いかにして審査会委員に受容され、知的作業の負荷を削減し、効率的な支援を実現するかというシステムの設計の検討が不可欠である。

最後に、認定審査には決定論的に判断される明確なロジックが存在するため、確率的な推論を行う LLM だけでなく、長年知見が蓄積されてきたルールベースの判断ロジックを組み込んだハイブリッドなシステムとして構築することで、判定の精度と信頼性を最大限に高めたより完成度の高いシステムとできることが期待される。

E. 結論

本研究課題は、非構造化テキストの解析という現代の AI 技術との親和性が高いタスクを含み、その実用化と将来的な社会実装は十分に期待で

きるものである。今後は、医療分野において重要である説明可能性と精度を両立させ、ユーザーである審査会委員に受容されるようなエッジ環境でのシステムを構築することが、実効性の鍵となると考えられる。

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

なし

H. 知的財産権の出願・登録状況

1. 特許取得

なし

2. 実用新案登録

なし

3. その他

なし

参考文献・参考資料

1. 総務省統計局. 「人口推計」. 政府統計の総合窓口(e-Stat). www.e-stat.go.jp, 2026 年 5 月閲覧.
2. 厚生労働省老健局介護保険課. 「介護保険事業状況報告の概要（平成 19 年 1 月暫定版）」. 厚生労働省 HP. www.mhlw.go.jp/topics/kaigo/osirase/jigyo/m07/dl/0701.pdf. 2026 年 5 月閲覧.
3. 厚生労働省老健局介護保険課. 「介護保険事業状況報告の概要（令和 7 年 12 月暫定版）」. 厚生労働省 HP. www.mhlw.go.jp/topics/kaigo/osirase/jigyo/m25/dl/2512a.pdf. 2026 年 5 月閲覧.
4. 厚生労働省老健局介護保険課. 「第 3 回介護 DB オープンデータ」. 厚生労働省 HP. www.mhlw.go.jp/stf/seisakunitsuite/bunya/huku-shi_kaigo/kaigo_koureisha/nintei/index_00013.html. 2026 年 5 月閲覧.
5. 厚生労働省老健局介護保険課. 「認定審査期間等の公表・認定審査期間等（令和 6 年度分）」. 厚生労働省 HP. www.mhlw.go.jp/content/12300000/001683967.xlsx. 2026 年 5 月閲覧.
6. Shortliffe, E. H., Axline, S. G., Buchanan, B. G., Merigan, T. C., & Cohen, S. N. (1973). An artificial intelligence program to advise physicians regarding antimicrobial therapy. *Computers and biomedical research, an international journal*, 6(6), 544–560. [https://doi.org/10.1016/0010-4809\(73\)90029-3](https://doi.org/10.1016/0010-4809(73)90029-3)
7. Shortliffe E. H. (1977). Mycin: A Knowledge-Based Computer Program Applied to Infectious Diseases. *Proceedings of the Annual Symposium on Computer Application in Medical Care*, 66–69.
8. Colby, K. M., Weber, S., & Hilf, F. D. (1971). Artificial paranoia. *Artificial intelligence*, 2(1), 1-25.
9. Isner, D. W. (1971, November). An inferential processor for interacting with biomedical data using restricted natural language. In *Proceedings of the May 16-18, 1972, spring joint computer conference* (pp. 1107-1124).
10. Vapnik, V. N. (1963). Pattern recognition using generalized portrait method. *Automation and remote control*, 24(6), 774-780.
11. Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992, July). A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory* (pp. 144-152).
12. Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3), 273-297.
13. Furey, T. S., Cristianini, N., Duffy, N., Bednarski, D. W., Schummer, M., & Haussler, D. (2000). Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics (Oxford, England)*,

- 16(10), 906–914. <https://doi.org/10.1093/bioinformatics/16.10.906>
14. Breiman, L., Friedman, J. H., Olshen, R., & Stone, C. J. (1984). Classification and Regression Trees.
15. Breiman, L. Random Forests. *Machine Learning* 45, 5–32 (2001).
<https://doi.org/10.1023/A:1010933404324>
16. Gunther, E. C., Stone, D. J., Gerwien, R. W., Bento, P., & Heyes, M. P. (2003). Prediction of clinical drug efficacy by classification of drug-induced genomic expression profiles in vitro. *Proceedings of the National Academy of Sciences of the United States of America*, 100(16), 9608–9613.
<https://doi.org/10.1073/pnas.1632587100>
17. Friedman, J. H. (2001): *Greedy function approximation: A gradient boosting machine*.
<https://www.jstor.org/stable/2699986>
18. Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785–794). <https://doi.org/10.1145/2939672.2939785>
19. Larkey, L. S., & Croft, W. B. (1996, August). Combining classifiers in text categorization. In *Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 289–297).
<https://doi.org/10.1145/243199.243276>
20. Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6), 386–408. <https://doi.org/10.1037/h0042519>
21. Hedén, B., Ohlin, H., Rittner, R., & Edenbrandt, L. (1997). Acute myocardial infarction detected in the 12-lead ECG by artificial neural networks. *Circulation*, 96(6), 1798–1802.
<https://doi.org/10.1161/01.cir.96.6.1798>
22. Cireşan, D. C., Giusti, A., Gambardella, L. M., & Schmidhuber, J. (2013). Mitosis detection in breast cancer histology images with deep neural networks. In *International conference on medical image computing and computer-assisted intervention* (pp. 411–418). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-40763-5_51
23. Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., & Kang, J. (2020). BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics (Oxford, England)*, 36(4), 1234–1240.
<https://doi.org/10.1093/bioinformatics/btz682>
24. Iqbal, U., Tanweer, A., Rahmanti, A. R., Greenfield, D., Lee, L. T., & Li, Y. J. (2025). Impact of large language model (ChatGPT) in healthcare: an umbrella review and evidence synthesis. *Journal of biomedical science*, 32(1), 45.
<https://doi.org/10.1186/s12929-025-01131-z>
25. Balan, R., & Gumpel, T. P. (2025). ChatGPT Clinical Use in Mental Health Care: Scoping Review of Empirical Evidence. *JMIR mental health*, 12, e81204. <https://doi.org/10.2196/81204>
26. Shool, S., Adimi, S., Saboori Amleshi, R., Bitaraf, E., Golpira, R., & Tara, M. (2025). A systematic review of large language model (LLM) evaluations in clinical medicine. *BMC medical informatics and decision making*, 25(1), 117.
<https://doi.org/10.1186/s12911-025-02954-4>
27. Bobkov, A., Cheng, F., Xu, J., Bobkova, T., Deng, F., He, J., Jiang, X., Khuzin, D., & Kang, Z. (2025). Telepsychiatry and Artificial Intelligence: A Structured Review of Emerging Approaches to Accessible Psychiatric Care. *Healthcare (Basel, Switzerland)*, 13(11), 1348.
<https://doi.org/10.3390/healthcare13111348>
28. Mansoor, I., Abdullah, M., Rizwan, M. D., & Fraz, M. M. (2025). Reasoning with large language models in medicine: a systematic review of techniques, challenges and clinical integration. *Health information science and systems*, 14(1), 6.
<https://doi.org/10.1007/s13755-025-00403-0>

29. Martinez-Ortigosa, A., Martinez-Granados, A., Gil-Hernández, E., Rodriguez-Arrastia, M., Romero-Padilla, C., & Roman, P. (2023). Applications of Artificial Intelligence in Nursing Care: A Systematic Review. *Journal of nursing management*, 2023, 3219127. <https://doi.org/10.1155/2023/3219127>
30. Oh, N., Jung, K. H., & Choi, G. S. (2026). Physical AI goes to the operating room: are we ready for the Surgical Data Factory?. *Annals of surgical treatment and research*, 110(3), 135–143. <https://doi.org/10.4174/astr.2026.110.3.135>
31. Mir, B. A., Nishwa, D. E., & Lee, S. W. (2026). Embodied Artificial Intelligence in Healthcare: A Systematic Review of Robotic Perception, Decision-Making, and Clinical Impact. *Healthcare (Basel, Switzerland)*, 14(5), 572. <https://doi.org/10.3390/healthcare14050572>
32. Wang, D., & Zhang, S. (2024). Large language models in medical and healthcare fields: applications, advances, and challenges. *Artificial intelligence review*, 57(11), 299. <https://doi.org/10.1007/s10462-024-10921-0>
33. Maity, S., & Saikia, M. J. (2025). Large Language Models in Healthcare and Medical Applications: A Review. *Bioengineering (Basel, Switzerland)*, 12(6), 631. <https://doi.org/10.3390/bioengineering12060631>
34. Rouzrokh, P., Khosravi, B., Faghani, S., Moassefi, M., Shariatnia, M. M., Rouzrokh, P., & Erickson, B. (2025). A Current Review of Generative AI in Medicine: Core Concepts, Applications, and Current Limitations. *Current reviews in musculoskeletal medicine*, 18(7), 246–266. <https://doi.org/10.1007/s12178-025-09961-y>
35. Fukui, Y., Kawata, Y., Kobashi, K., Nagatani, Y., & Iguchi, H. (2025). Evaluation of a retrieval-augmented generation system using a Japanese Institutional Nuclear Medicine Manual and large language model-automated scoring. *Radiological physics and technology*, 18(3), 861–876. <https://doi.org/10.1007/s12194-025-00941-y>
36. Collaco, B. G., Srinivasagam, P., Gomez-Cabello, C. A., Haider, S. A., Genovese, A., Wood, N. G., Bagaria, S., Lifson, M. A., & Forte, A. J. (2026). Integrating Fine-Tuning and Retrieval-Augmented Generation for Healthcare AI Systems: A Scoping Review. *Bioengineering (Basel, Switzerland)*, 13(2), 225. <https://doi.org/10.3390/bioengineering13020225>
37. Crowe, B., Shah, S., Teng, D., Ma, S. P., DeCamp, M., Rosenberg, E. I., Rodriguez, J. A., Collins, B. X., Huber, K., Karches, K., Zucker, S., Kim, E. J., Rotenstein, L., Rodman, A., Jones, D., Richman, I. B., Henry, T. L., Somlo, D., Pitts, S. I., Chen, J. H., ... & Mishuris, R. G. (2025). Recommendations for Clinicians, Technologists, and Healthcare Organizations on the Use of Generative Artificial Intelligence in Medicine: A Position Statement from the Society of General Internal Medicine. *Journal of general internal medicine*, 40(3), 694–702. <https://doi.org/10.1007/s11606-024-09102-0>
38. Dell’Acqua, F., McFowland III, E., Mollick, E., Lifshitz, H., Kellogg, K. C., Rajendran, S., ... & Lakhani, K. R. (2026). Navigating the jagged technological frontier: Field experimental evidence of the effects of artificial intelligence on knowledge worker productivity and quality. *Organization Science*, 37(2), 403–423. <https://doi.org/10.1287/orsc.2025.21838>