

生成AIがもたらす 医療変革

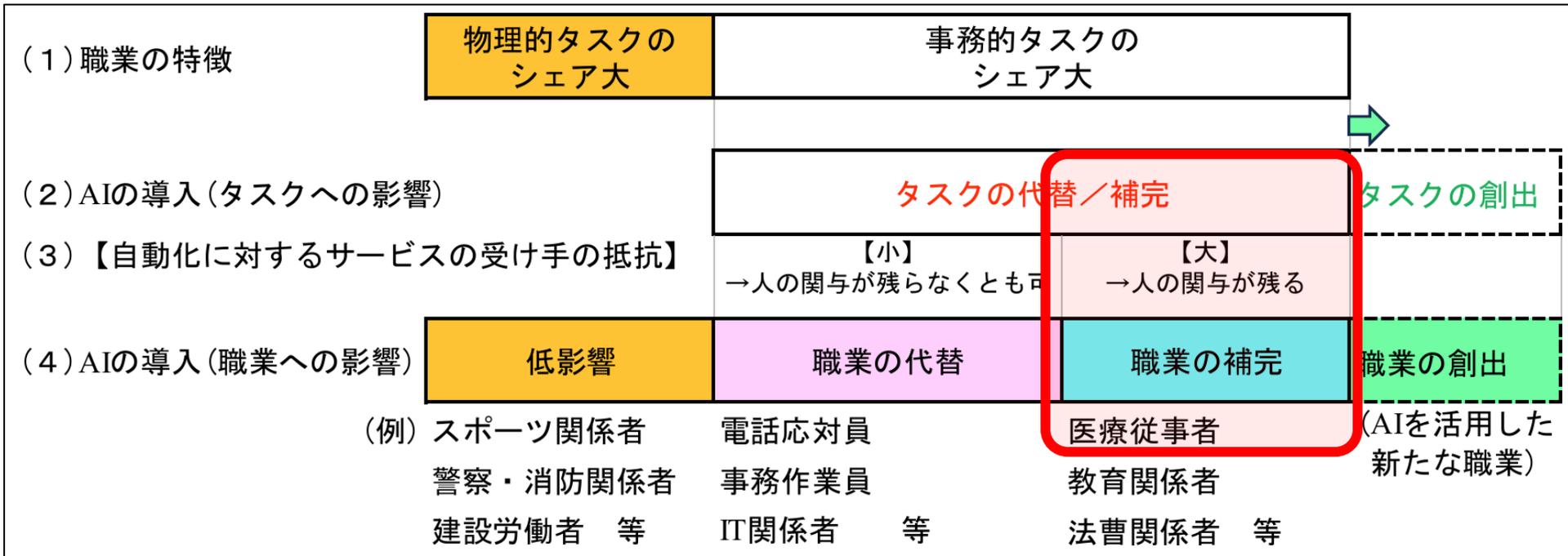
医療分野における生成AIの台頭は、医療従事者の
業務効率化と研究推進に革命をもたらさうる

最新の動向と実践例を通じて、その可能性と課題を探る



岡山大学学術研究院医歯薬学域 長谷井嬢

生成AIが与える労働市場への影響



(2) 影響が大きく、補完性が高い職業	
1	歯科医師
2	医師
3	裁判官
4	弁護士
5	教員 (高等学校)
6	建築施工管理技術者
7	研究者 (自然科学系)
8	助産師
9	看護師
10	作業療法士
11	保育士
12	会社管理職員
13	教員 (中学校)
14	教員 (高等専門学校)
15	会社役員
16	薬剤師
17	聖職者
18	言語聴覚士
19	記者
20	通訳家・翻訳家

- 医療者にもAIの基本的な知識や活用スキルが求められる可能性がある
- AIの出力結果を批判的に考察する能力が必要
- 人間にしかできないことへの集中: 患者とのコミュニケーションや共感複雑な判断を要する診断・治療方針の決定倫理的な判断
- AIとの協働による医療の質向上

世界経済の潮流

2024年 I

AIで変わる労働市場

令和6年7月

内閣府

政策統括官 (経済財政分析担当)



医師の働き方改革の現状と課題

1

時間外労働規制

2024年から、A水準(年960時間)、B水準(年1,860時間)の上限規制

2

大学病院の課題

研究・教育・高度医療の両立
診療時間と研究時間の確保
困難

3

業務負担の増大

膨大な文書作成業務と医療情報の増加により、知識アップデート
負担が大きくなっている

生成AIによるソリューション提案

1

業務効率化

文書作成支援や文献レビュー・要約の効率化により、研究時間を創出

2

データ解析支援

データ解析の自動化・効率化により、研究プロセスを加速

3

研究プロセス革新

プロトコル作成支援や統計解析の支援、論文作成補助により研究の質を向上



A Survey of Large Language Models in Medicine: Progress, Application, and Challenge

生成AIの進化と医療における動向

A Survey of Large Language Models in Medicine: Progress, Application, and Challenge

Hongjian Zhou^{1*}, Fenglin Liu^{1,2,†}, Boyang Gu^{2*}, Xinyu Zou^{1*}, Jinfa Huang^{4*},
Jinge Wu⁵, Yiru Li⁶, Sam S. Chen⁷, Peilin Zhou⁸, Junling Liu⁹, Yining Hua¹⁰,
Chengfeng Mao¹¹, Chenyu You¹², Xian Wu¹³, Yefeng Zheng¹⁴, Lei Clifton¹,
Zheng Li^{14,†}, Jiebo Luo^{4,†}, David A. Clifton^{1,15,†}

* Core Contributors, ordered by a coin toss. † Corresponding Authors.

¹ University of Oxford, ² Imperial College London, ³ University of Waterloo,

⁴ University of Rochester, ⁵ University College London, ⁶ Western University,

⁷ University of Georgia, ⁸ Hong Kong University of Science and Technology (Guangzhou),

⁹ Alibaba, ¹⁰ Harvard T.H. Chan School of Public Health, ¹¹ Massachusetts Institute of Technology,

¹² Yale University, ¹³ Tencent, ¹⁴ Amazon, ¹⁵ Oxford-Suzhou Centre for Advanced Research

{hongjian.zhou@cs.ox.ac.uk, fenglin.liu@eng.ox.ac.uk, anzzhe@amazon.com, {jhuang90@ur.jluo@cs.rochester.edu

Abstract

Large language models (LLMs), such as ChatGPT, have received substantial attention due to their capabilities for understanding and generating human language. While there has been a burgeoning trend in research focusing on the employment of LLMs in supporting different medical tasks (e.g., enhancing clinical diagnostics, and providing medical education), a comprehensive review of these efforts, particularly their development, practical applications, and outcomes in medicine, remains scarce. Therefore, this review aims to provide a detailed overview of the development and deployment of LLMs in medicine, including the challenges and opportunities they face. In terms of development, we provide a detailed introduction to the principles of existing medical LLMs, including their basic model structures, number of parameters, and sources and scales of data used for model development. It serves as a guide for practitioners in developing medical LLMs tailored to their specific needs. In terms of deployment, we offer a comparison of the performance of different LLMs across various medical tasks, and further compare them with state-of-the-art lightweight models, aiming to provide a clear understanding of the distinct advantages and limitations of LLMs in medicine. Overall, in this review, we address the following study questions: 1) What are the practices for developing medical LLMs? 2) How to measure the medical task performance of LLMs in a medical setting? 3) How have medical LLMs been employed in real-world practice? 4) What challenges arise from the use of medical LLMs? and 5) How to more effectively develop and deploy medical LLMs? By answering these questions, this review aims to provide insights into the opportunities and challenges of LLMs in medicine and serve as a practical resource for constructing effective medical LLMs. We also maintain a regularly updated list of practical guides on medical LLMs at <https://github.com/AI-in-Health/MedLLMPracticalGuide>.

Preprint. Version 4.

Submission history

From: Fenglin Liu [[view email](#)]

[v1] Thu, 9 Nov 2023 02:55:58 UTC (48 KB)

[v2] Mon, 11 Dec 2023 18:10:20 UTC (199 KB)

[v3] Fri, 2 Feb 2024 06:48:24 UTC (417 KB)

[v4] Sun, 3 Mar 2024 01:15:36 UTC (652 KB)

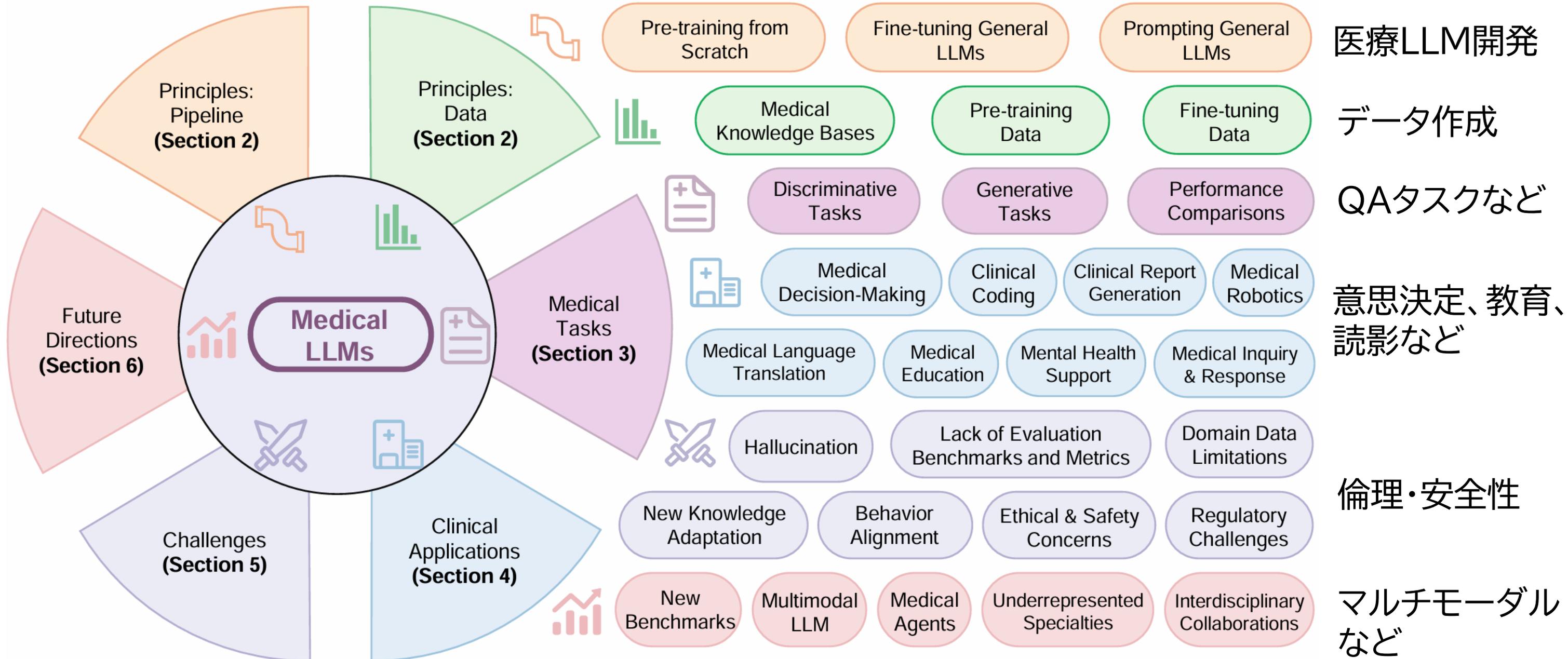
[v5] Wed, 15 May 2024 13:38:45 UTC (1,049 KB)

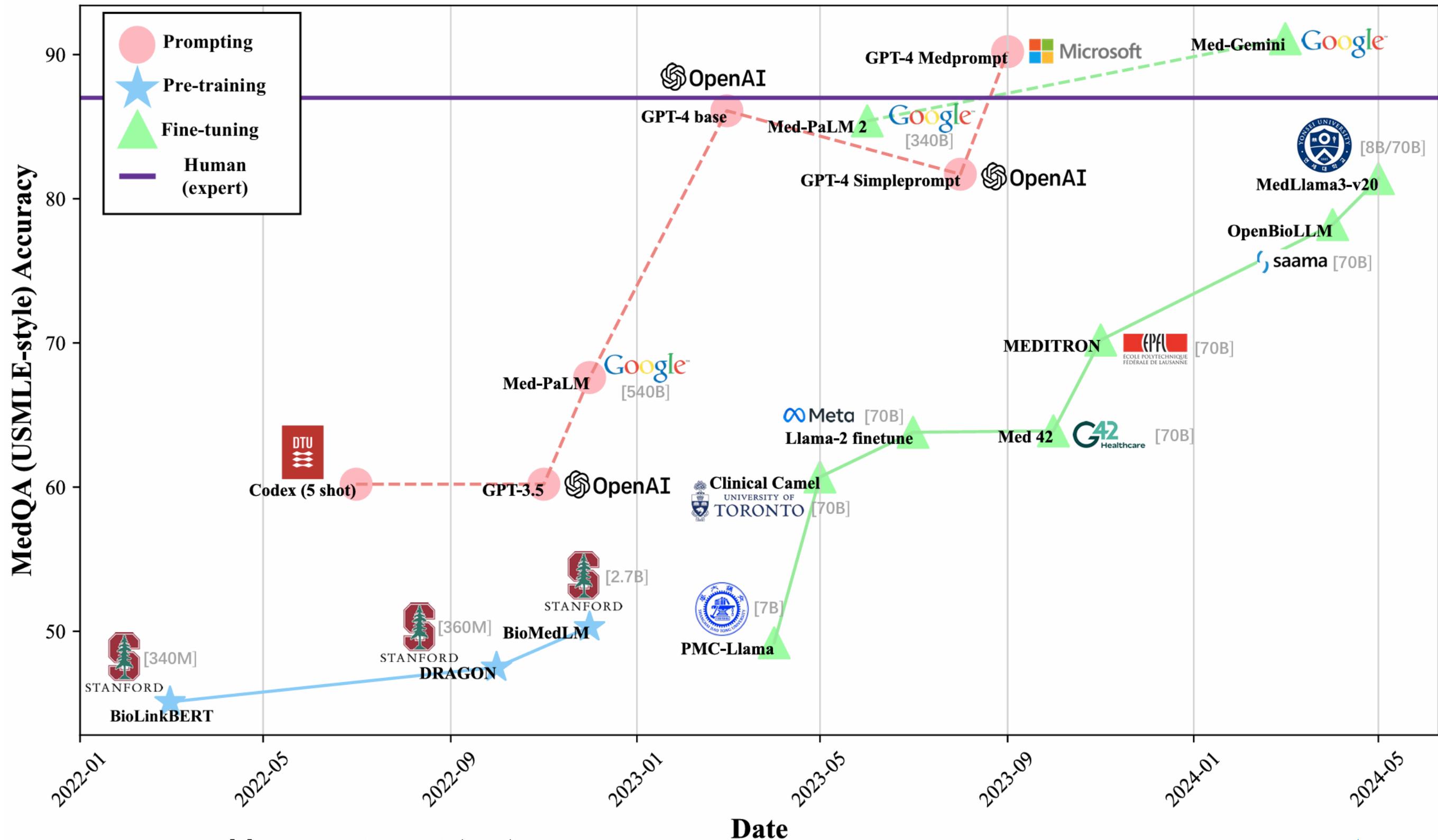
[v6] Wed, 10 Jul 2024 22:10:32 UTC (1,168 KB)

[v7] Mon, 22 Jul 2024 12:16:30 UTC (1,168 KB)

[2] Hongjian Zhou et al. (2023), "[A Survey of Large Language Models in Medicine: Progress, Application, and Challenge \(arxiv.org\)](#)" より引用

生成AIの医療応用の現在地

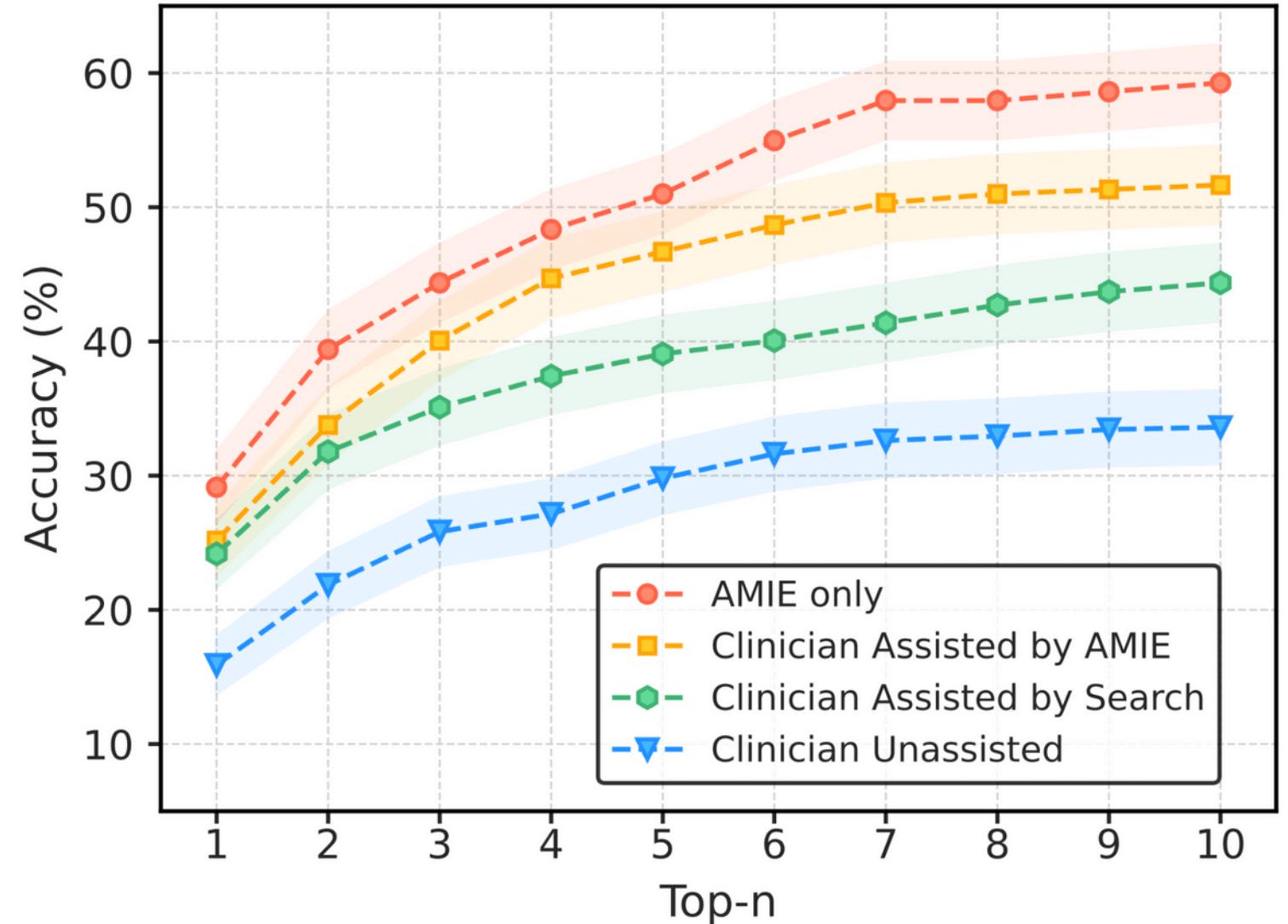
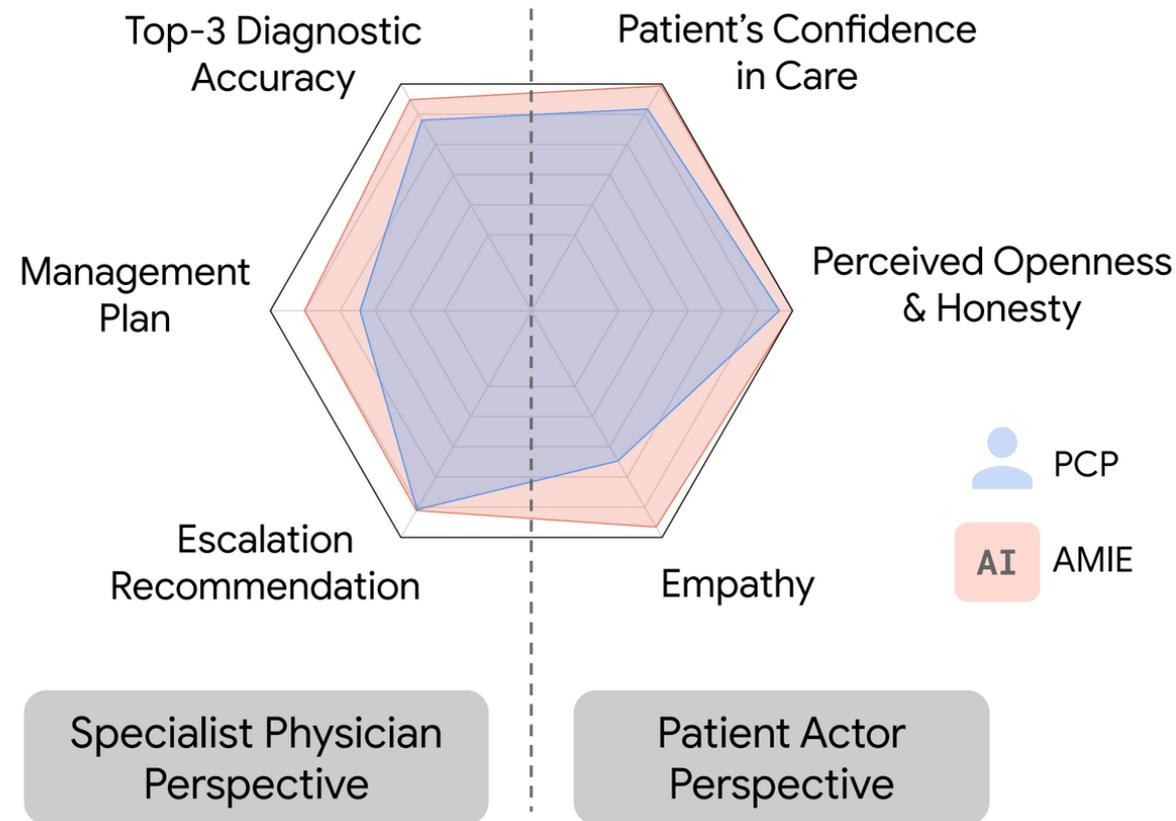




[2] Hongjian Zhou et al. (2023), “A Survey of Large Language Models in Medicine: Progress, Application, and Challenge (arxiv.org)” より引用

nature

Google AI has better bedside manner than human doctors – and makes better diagnoses

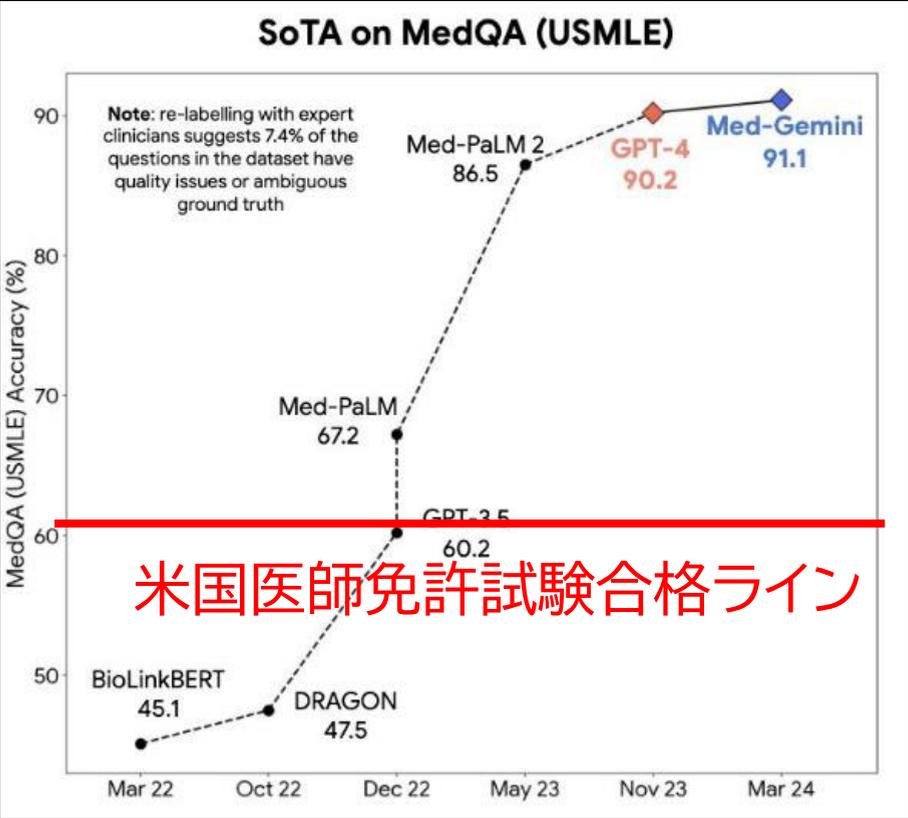


「診断の正確さ」「治療に対する信頼感」「医師の誠実さ」「医師からの共感」「指示の的確さ」「患者の健康管理」の6つ基準全てで、人間の医師の診断精度に匹敵またはそれを上回る結果

[3] Mariana Lenharo. (2024), "[Google AI has better bedside manner than human doctors — and makes better diagnoses](#)" より引用

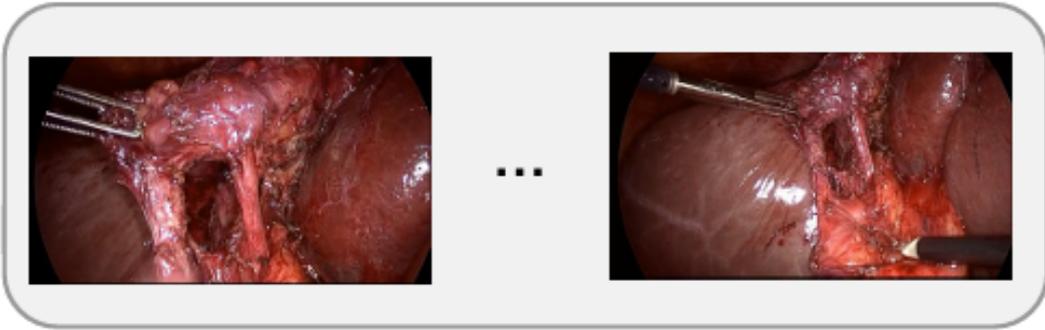
Googleが医療版Gemini(Med-Gemini)を発表

2024/05/01



(a) Video dialogue example

Hello, I am a medical student analyzing this surgical procedure video for my studies. Which surgical procedure is shown in this video?



The surgical procedure shown in this video is a laparoscopic cholecystectomy - the surgical removal of the gallbladder. This is a common procedure and is often performed when gallstones cause pain or complications, such as inflammation of the gallbladder (cholecystitis).

Which tools and how many are being used by the surgeon?

Image Classification (MIMIC-CXR 13-condition classification)

Visual input

Instruction
You are a helpful radiology assistant. The following are questions about findings in chest X-ray in the frontal view. Identify if a specific type of abnormality is shown in the X-ray.
Given the <VIEW> X-ray image,
Question: Which of the following abnormalities are indicated by the image? (A) Atelectasis (B) Cardiomegaly (C) Consolidation (D) Edema (E) Enlarged Cardiomeastinum (F) Fracture (G) Lung Lesion (H) Lung Opacity (I) Pleural Effusion (J) Pleural Other (K) Pneumonia (L) Pneumothorax (M) Support Devices

Response
(A)

A) 無気肺、B) 心肥大、C) 浸潤影、D) 肺水腫...

Q:「この動画は、何の手術ですか？」
Med-Gemini:「これは、腹腔鏡下胆嚢摘出術で、胆嚢管と胆嚢動脈を剥離して切除しているところです。」

[4] Khaled Saab et al. (2024), "[2404.18416] Capabilities of Gemini Models in Medicine" より引用

医療における検索拡張生成 (RAG: Retrieval-Augmented Generation)

5.4 New Knowledge Adaptation

LLMs are trained on extensive data to learn knowledge. Once trained, it is expensive and inefficient to inject new knowledge into an LLM through re-training. However, it is sometimes necessary to update the knowledge of the LLM, for example, on a new adverse effect of a medication or a novel disease. Two problems occur during such knowledge updates. The first problem is how to make LLMs appropriately ‘forget’ the old knowledge, as it is almost impossible to remove all ‘old knowledge’ from the training data, and the discrepancy between new and old knowledge can cause unintended association and bias [56]. The second problem is the timeliness of the additional knowledge - how do we ensure the model is updated in real-time [210]? Both problems pose significant barriers to using LLMs in medical fields, where accurate and timely updates of medical knowledge are crucial in real-world implementations.

Potential Solutions Current solutions to knowledge adaptation can be categorized into model editing and retrieval-augmented generation. Model editing [211] alters the knowledge of the model by modifying its parameters. However, this method does not generalize well, with their effectiveness varying across different model architectures. In contrast, retrieval-augmented generation provides external knowledge sources as prompts during model inference; for example, Lewis et al. [212] enabled model knowledge updates by updating the model’s external knowledge memory.

医療情報は薬剤の新しい副作用や、適応追加、
要領変更など、タイムリーな変更がある



RAGでの対応は有意義

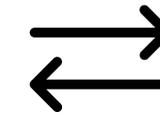
予め用意したテキスト情報から検索し
回答を生成

Pythonプログラム実行

文書データベース



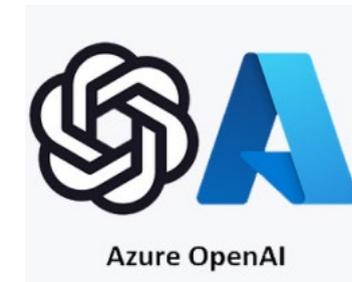
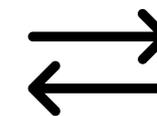
Azure
Functions



Database



LangChain



岡山大学での取り組み：実践例

介護保険回答AI

医療従事者から患者への説明業務の効率化

64歳です
介護保険は利用出来ますか 12:02

介護保険は65歳以上の方が利用できる制度ですが、40歳から64歳の方で特定の疾病に該当する場合も利用できます。申請は介護保険課で行い、申請書には主治医の情報を記入する必要があります。申請後は認定調査がありますので、自宅や病院で調査を受けることになります。申請から認定が降りるまでには平均45日かかると言われています。介護保険の利用には年齢制限や申請手続きがありますので、詳細は介護保険課に相談してください。

骨軟部腫瘍AI, 小児股関節AI

専門領域特化型の言語AI 非専門医への情報提供

問 4-8-7
多発する骨腫瘍として誤っているのはどれか。
a. 骨軟骨腫 b. 線維性骨異形成 c. 内軟骨腫 d. Langerhans細胞組織球症 e. 類骨骨腫



質問に対する答えとして、多発する骨腫瘍として誤っているものを選ぶ必要があります。選択肢の中で、一般的に多発する骨腫瘍として誤っているのは「e. 類骨骨腫」です。類骨骨腫は通常、単発性であり、多発することは稀です。他の選択肢は多発する可能性がある腫瘍です。

午後 8:18

メンタルケアAI

ヒトには言えない悩み相談、ヒトでは十分でないカウンセリング時間を支援



既読
午後 2:53 針を何回もさされました



それは本当に痛かったですよね。何回も針を刺されるのは、心身ともに大変なストレスだと思います。どのくらいの頻度で検査を受けているんですか？

午後 2:53

医療分野での説明業務問題

- ✓医療文書の説明を受けられる時間は限られている
- ✓情報量が多く、その場で理解するのが難しい
- ✓あとから質問をしたくてもできない



既読
午後 4:34

50歳ですが、介護保険りようできますか？

2024年4月から患者利用可能



質問: 介護保険は50歳の人でも利用できますか？

回答: 介護保険は65歳以上で介護保険料を収めている方が利用できる制度です。また、40歳から64歳の方で、16の特定疾病に該当する方も利用ができます。ですので、50歳の方でも介護保険を利用することができます。

午後 4:34

既読
23:29

介護認定を受けるとどんなメリットがあるのですか？

介護認定を受けると、以下のようなメリットがあります：

1. 介護サービスの利用が可能になります。介護認定を受けることで、適切なサービスプランを作成し、必要なサービスを受けることができます。
2. 自己負担額が割引されます。介護保険の自己負担額は、収入やサービスの種類によって異なりますが、一般的にはサービス費用の1割から3割です。
3. 上限付きでサービスを利用できます。介護保険では、利用者の認定レベルに応じた上限が設定されており、必要なサービスを適切に利用できます。
4. 異議申し立てが可能です。介護認定結果に異議がある場合は、市区町村に異議申し立てを行うことがで



既読
20:40

介護サービスを利用しなくても保険料は払わなくてはいけませんか？



介護サービスを利用しなくても、保険料は払わなくてはいけません。介護保険料は、国民皆保険制度の一環として、年収に応じて徴収されます。介護保険料は、介護サービスの利用に関係なく支払われるため、利用しなくても免除されることはありません。ただし、所得が低い場合や特定の条件を満たす場合には、介護保険料の免除申請が可能です。具体的な条件は市区町村によって異なるので、詳細は役所にお問い合わせください。

20:40

既読
20:43

家を売りました。介護保険は上がりますか？



介護保険は、家を売っても上がることはありません。介護保険は、65歳以上の方や40歳から64歳の特定疾病に該当する方が利用できる制度です。介護保険の申請は介護保険課で手続きができます。申請時には介護保険証を持参する必要があります。申請が終わったら、認定調査があります。自宅または入院中であれば病院で調査を受けます。介護保険の利用には、介護認定を受けてサービスプランを作成し、サービス提供者を選ぶ必要があります。

20:43

An astronaut in a white spacesuit with a large life-support backpack is walking on the lunar surface. To the right, a lunar lander is parked. The Earth is visible in the background as a large, bright, curved horizon against the black sky of space. The lunar surface is covered in craters and rocks.

□ 大学病院内での説明業務にAIを導入した**国内初**の試み

□ 今後の医療現場での生成AIの活用において**極めて重要な第一歩**

骨軟部腫瘍特化型LLM

骨盤(仙骨を含む)が第一の好発部位である悪性骨腫瘍はどれか。3つ選べ。

- a. 脊索腫
- b. 軟骨肉腫
- c. 通常型骨肉腫
- d. 傍骨性骨肉腫
- e. Ewing肉腫



質問に対する答えは以下の3つです：

- a. 脊索腫
- b. 軟骨肉腫
- e. Ewing肉腫

これらの悪性骨腫瘍は、骨盤（仙骨を含む）を第一の好発部位とすることが多いです。

多発する骨腫瘍として誤っているのはどれか。

- a. 骨軟骨腫
- b. 線維性骨異形成
- c. 内軟骨腫
- d. Langerhans細胞組織球症
- e. 類骨骨腫

正解	モデル名	正誤	回答
e	骨軟部腫瘍特化型	○	e
	GPT-4o	×	a

既読
午後 8:18

多発する骨腫瘍として誤っているのはどれか。
a. 骨軟骨腫 b. 線維性骨異形成 c. 内軟骨腫 d. Langerhans細胞組織球症 e. 類骨骨腫



質問に対する答えとして、多発する骨腫瘍として誤っているものを選ぶ必要があります。選択肢の中で、一般的に多発する骨腫瘍として誤っているのは「e. 類骨骨腫」です。類骨骨腫は通常、単発性であり、多発することは稀です。他の選択肢は多発する可能性がある腫瘍です。

午後 8:18

骨軟部腫瘍特化型



この問題は「多発する骨腫瘍として誤っているもの」を問っています。

正解は a) 骨軟骨腫です。

理由：

- 骨軟骨腫は通常、単発で発生します。稀に多発する場合がありますが、それは遺伝性の多発性骨軟骨腫症の場合に限られ、一般的には単発で見られます。

GPT-4o



股関節領域への応用を試す

第36回 整形外科専門医試験

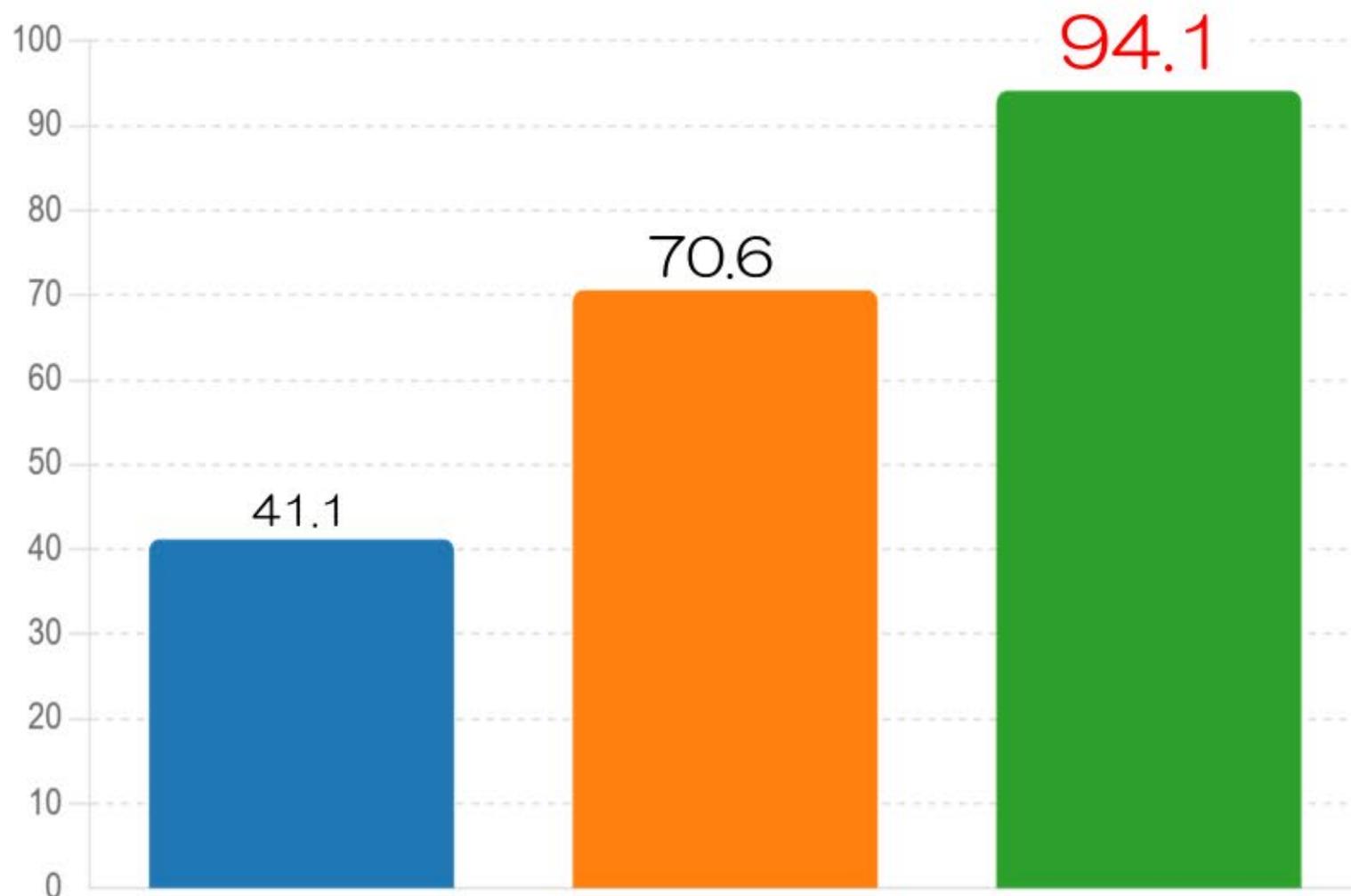
問 35 生後9日目の女児. 第1子として在胎 39週 3 日, 頭位正常分娩で出生した. 新生児健診で左股関節の開排制限を指摘された. 股関節の開排は右が 80° , 左が 60° , Allis徴候は陽性(左膝が右膝より低い)で, クリック(整復感)は認めなかった. Graf 法による股関節超音波検査画像ではtype IIに該当する. 初期対応で正しいのはどれか.

- a. 徒手整復
- b. 観血的整復
- c. 頭上方向牽引
- d. 抱き方など育児指導
- e. Pavlik 法(Riemenbügel 法)

第36回専門医試験 + 卒後研修Q&A
小児股関節・大腿骨頭壊死領域 17問
GPT-4: 7問正解(41.2%)

正解	モデル名	正誤	回答
d	GPT-4	×	e

モデル名	問題数	正解数	正解率(%)
GPT-4	17	7	41.1
小児股関節特化型	17	12	70.6
小児股関節特化型 (プロンプト調整済)	17	16	94.1



第36回 合格者平均：65.3点

GPT-4 小児股関節特化型 小児股関節特化型
プロンプト調整済



誰しものが専門知識を
自在に使用できる時代



★ 小児股関節説明AI

既読
午後 5:29

生後1週間の私の子供です、股関節が脱臼していると言われました。心配です。どんな病気なんですか？



股関節脱臼（先天性股関節脱臼、DDH）は、赤ちゃんの股関節が正常に発達しない状態を指します。これは、股関節が不安定で、容易に脱臼することが特徴です。遺伝的要因や胎内環境が影響することがあり、家族内での発生も報告されています。

治療は早期に行うことが重要で、通常は装具を使用して股関節を正しい位置に保つ方法が取られます。重度の場合は手術が必要になることもあります。心配な場合は、専門医に相談して適切な治療を受けることが大切です。

午後 5:30



実装における課題

- ① データセキュリティ
匿名化技術の導入と
厳格なアクセス管理
患者情報保護
- ② AIリテラシー向上
医療スタッフの
AI活用スキル向上
- ③ 倫理的配慮
AI倫理委員会を設置し、公平性と透明性を確保



今後の展望とまとめ

研究環境の変革

生成AIにより、大学病院の研究環境が劇的に改善される可能性

効率化の重要性

人海戦術や根性論からの脱却が必要
AIを活用した効率的な医療研究が求められている

次世代医療者の環境整備

デジタルネイティブ世代の医師が活躍できる、AI統合型の医療環境を整備することが重要