

2026/03/06

PMDAシンポジウム「生成AIの医療活用の最前線」

# 生成AIは医療現場でどこまで活用されているのか！？

国立がん研究センター研究所・主任研究員  
国立情報学研究所・特任准教授

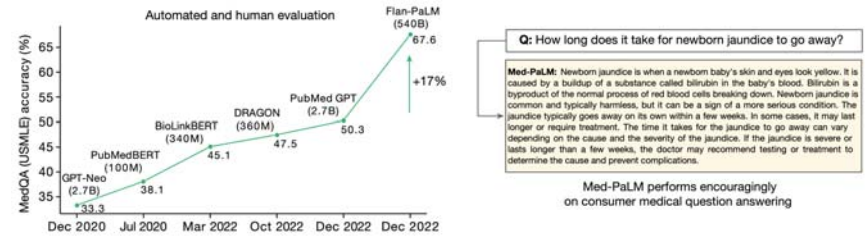
小林 和馬



## 医療分野における生成AIの可能性

- 従来、人間の医師にとってのコア・コンピテンシーと考えられていた様々な臨床能力において、近年、大規模言語モデル（LLM）が人間と同等のパフォーマンスを発揮することが報告されるようになってきた。
  - 医師国家試験（Singhal K., et al. Nature. 2023）
  - 診断困難例に対する診断（Alexander V.E., et al. NEJM AI. 2023）
  - 臨床技能・態度など（Tao T., et al. Nature. 2025）

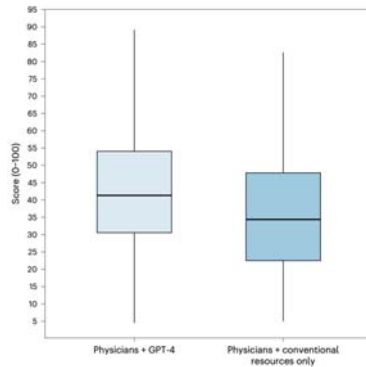
LLMが米国医師国家試験の合格水準に達する



Singhal K., et al. Large language models encode clinical knowledge. Nature, 2023

## LLMを用いた医師における診断スコアの向上

- LLMを用いた医師は、LLMを用いなかった医師に比べて、複雑な臨床シナリオにおける診断スコアが向上する。

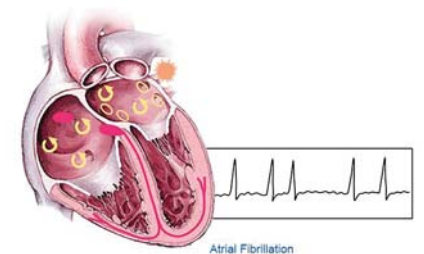


Ethan G., et al. GPT-4 assistance for improvement of physician performance on patient care tasks: a randomized controlled trial. Nature Medicine, 2025

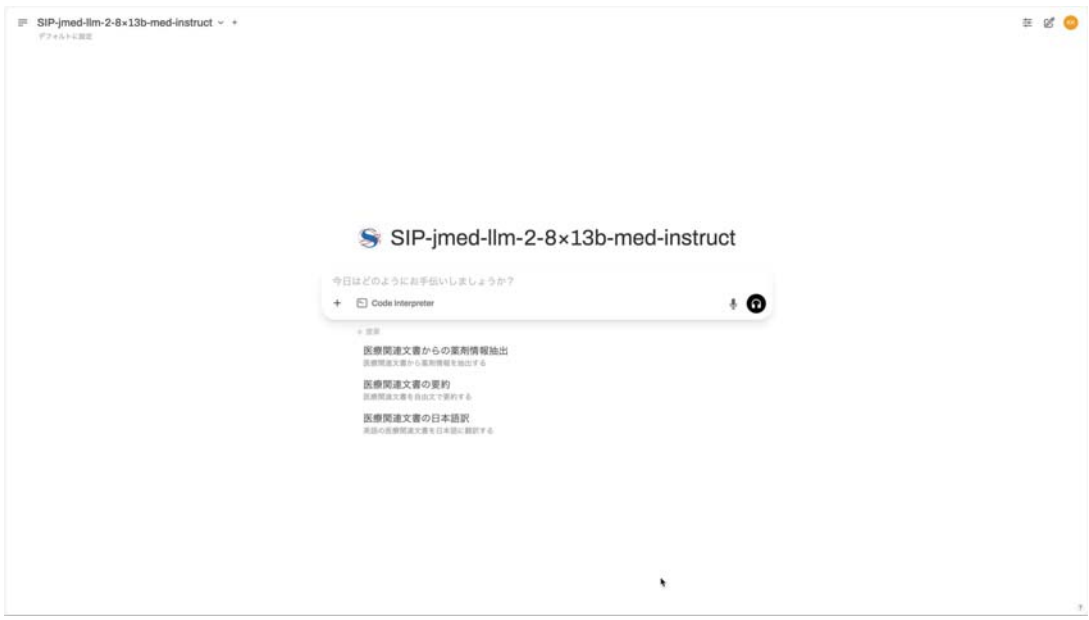
## 72歳女性。胆嚢穿孔の術後に心房細動を発症。

- 72歳の女性患者。GERD、2型糖尿病、コントロール不良の高血圧の既往歴があり、発熱、悪寒、腹痛を主訴として受診。腹部CT検査で胆嚢穿孔性胆嚢炎と腹膜炎が認められた。広域スペクトラム抗生物質が投与され、緊急開腹胆嚢摘出術が行われた。手術は順調に終了し、直後の合併症はなかった。術後1日目、テレメトリーで無症状の心房細動が短時間検出されたため、看護スタッフから呼び出された。テレメトリーデータの詳しい確認によると、過去24時間で心房細動が3回発生し、それぞれ15～20秒間持続していた。胆嚢炎と腹膜炎の観点からは、臨床状態は改善しており、手術後発熱や悪寒はなく、液体食を摂取可能です。血圧と心拍数は現在正常です。検査所見では白血球数の減少傾向が認められます。術前心電図と現在の心電図ともに、正常洞調律で、重大な異常は認められません。心エコー検査の既往はありません。現在、抗凝固薬や抗血小板薬は投与されていません。
- 質問: テレメトリーモニタリングを継続する以外に、現在、追加のモニタリング、検査、または治療を推奨しますか？もしそうなら、何ですか？

胆嚢穿孔にて入院・緊急手術となった  
72歳女性の術後に心房細動が認められた。  
今は落ち着いているが、  
追加の検査や治療は必要か？



<https://www.hopkinsmedicine.org/health/conditions-and-diseases/atrial-fibrillation> 画像引用



## 症例に対する適切な分析

SIP-jmed-llm-2-8x13b-med-instruct
6

72歳の女性患者で、GERD、2型糖尿病、コントロール不良の高血圧の既往歴があり、発熱、悪寒、腰痛を主訴として受診し、腹部CT検査で胆嚢穿孔性胆嚢炎と膵臓炎が確認されたため、広域スペクトラム抗生物質投与と緊急胆嚢摘出術が行われました。術後1日目に無症状の心房細動がテレメトリーで検出され、過去24時間で15~20秒間持続する心房細動が3回認められましたが、臨床状態は改善し、発熱や悪寒はなく、液体食摂取可能で、血圧と心拍数は正常、白血球数は減少傾向にあり、術前・現在の心電図は正常律で重大な異常はありませんでした。心エコー検査の既往はなく、抗凝固薬や抗血小板薬も投与されていません。

この患者に対して、テレメトリーモニタリングを継続することに加え、追加のモニタリング、検査、または治療を推奨するかどうかを検討します。

<b>追加のモニタリング、検査、または治療の必要性</b>	<b>患者の状態に関する正確な要約</b>
<b>1. 心房細動の持続時間と頻度の評価</b> - 現在の心房細動は短時間（15~20秒）で無症状ですが、頻度が増加する場合や持続時間が長くなる場合には、より詳細な評価が必要です。 - 24時間ホルター心電図や長時間のテレメトリーモニタリングを検討し、心房細動のパターンをさらに評価します。	<b>心房細動の正確な診断に向けた追加モニタリング</b>
<b>2. 心エコー検査の実施</b> - 心房細動の原因として心機能や弁膜症の評価が必要です。特に、左心室の拡大や心筋症の有無を確認するために、経胸壁心エコー検査を実施します。	<b>心エコーによる器質的因子の精査</b>
<b>3. 抗凝固療法の検討</b> - 無症状の心房細動であっても、CHA <sub>2</sub> DS <sub>2</sub> -VAScスコアやCHA <sub>2</sub> DS <sub>2</sub> -VAScスコアを用いて血栓症リスクの評価を行います。 - 高リスク患者（例：CHA <sub>2</sub> DS <sub>2</sub> -VAScスコア≥2）では、抗凝固薬（例：DOACやワルファリン）の投与を検討します。	<b>脳梗塞予防のためのリスク評価</b>
<b>4. 電解質と甲状腺機能の評価</b> - 電解質異常（特にカリウム、マグネシウム、カルシウム）や甲状腺機能亢進症が心房細動の原因となることがあるため、血液検査を実施します。	<b>心房細動の誘引探索</b>
<b>5. 術後のストレスや炎症の評価</b> - 術後のストレスや炎症が心房細動の誘因となることがあるため、CRPや他の炎症マーカーをモニタリングします。	<b>周術期における影響因子の探索</b>
<b>結論</b> この患者に対しては、テレメトリーモニタリングを継続することに加え、心エコー検査を実施し、心房細動の原因を評価することが推奨されます。また、抗凝固療法の必要性を評価するため、血栓症リスクスコアを計算し、必要に応じて抗凝固薬を開始します。さらに、電解質や甲状腺機能の検査を行い、術後のストレスや炎症の影響をモニタリングします。これらの追加評価と治療により、心房細動の管理と血栓症の予防を図ります。	<b>推奨される検査と治療の要約</b>

## 脳梗塞発症リスクの適切な推定と推奨治療

SIP-jmed-llm-2-8x13b-med-instruct
7

この患者のCHA<sub>2</sub>DS<sub>2</sub>-VAScスコアを正確に計算してください。

CHA<sub>2</sub>DS<sub>2</sub>-VAScスコアは、心房細動患者の血栓症リスクを評価するためのスコアリングシステムです。このスコアは、以下の危険因子に基づいて計算されます：

**スコアシステムの正しい解説**

1. Congestive heart failure (心不全): 1点
2. Hypertension (高血圧): 1点
3. Age ≥75 years (年齢≥75歳): 2点
4. Diabetes mellitus (糖尿病): 1点
5. Stroke/TIA (脳卒中または一過性脳虚血発作の既往): 2点
6. Vascular disease (血管疾患): 1点
7. Age 65-74 years (年齢65-74歳): 1点
8. Sex category (女性): 1点

この患者の背景を考慮すると、以下ようになります：

- 年齢: 72歳 (65-74歳の範囲内)
- 性別: 女性
- 既往歴:
  - 2型糖尿病 (1点)
  - 高血圧 (1点)
  - 胆嚢穿孔性胆嚢炎 (血管疾患や心不全の記載なし)

**正確なスコア計算とリスク評価**

合計スコアは4点となります。このスコアに基づき、この患者は中等度の血栓症リスクを有すると判断されます。

<https://my.clevelandclinic.org/health/diseases/embolic-stroke> 画像引用

**Embolic Stroke**

心房細動により心臓にできた血栓が脳血管を詰まらせることで脳梗塞が発症することがある

**脳梗塞発症リスクの推定にもとづき、血栓塞栓症予防の治療導入を推奨**

## 日本の医療に特化した大規模言語モデルの研究開発

SIP「統合型ヘルスケアシステムの構築」(PD: 永井 良三先生)
8

- SIP「統合型ヘルスケアシステムの構築」(PD: 永井 良三先生)の一貫として、医療LLMの研究開発を実施。
  - 事業名：SIP第3期補正「統合型ヘルスケアシステムの構築における生成AIの活用」
  - 研究開発期間: 令和6年度
  - 参加組織: 24組織 (情報分野7組織、医療分野17組織)
  - 研究開発責任者: 相澤 彰子先生 (国立情報学研究所)
- 医療現場での社会実装を目標
  1. 医療の質の向上、診療・受診支援、医療業務の効率化
    - 医学知識を統合し、専門医レベルの臨床意思決定支援を実現する
    - 医学的な因果関係の理解に根ざした診療業務支援の実現
  2. 医療行政支援
    - 病名のICDへの自動コーディング
    - 国内標準用語集への自動正規化支援
  3. 研究開発支援・医学知識の発見
    - カルテ情報から重要な臨床的イベントを自動抽出するシステム開発
    - ゲノム医療モデルの開発による個別化医療の実現 など

## 日本の医療に特化したソブリンLLMの開発意義

9

- 国民の健康と生命を守る医療分野において、日本の文化・言語・制度に根ざした医療LLMを、自国の国力にもとづいて研究開発することの意義は大きい。
  - データ主権とプライバシー保護: 医療情報は究極の個人情報
  - 医療従事者や患者に対する信頼確保: 開発過程の透明性担保による高いアカウンタビリティ

### 日本人固有の健康問題や疾患リスク

- 醤油・味噌・漬物など、発酵食品に由来する食塩摂取量が多く、そのために高血圧や脳血管疾患のリスクが高いと言われている。
- 副作用の少ない分子標的抗がん剤として承認されたイレツサは、日本人において重篤な肺障害を引き起こし、社会問題となった。



### 日本固有の文化や医療制度・体系への適合

- 終末期医療では、宗教観、家族間、死生観などの文化的慣習が大きな影響を与える（安楽死の制度的否定など）。
- 漢方医学などの日本特有の医学体系や、欧米の標準治療とは異なる治療戦略もある（直腸がんの側方リンパ節郭清など）。

## GPT-5は日本の医療をどれくらい理解しているのか？

10

遺伝性血管性浮腫の治療として、地方の小規模病院でも出来る救急外来での点滴治療を教えてください。

### GPT-5の回答

： (中略)

優先度	治療法	投与経路	可否 (地方病院)	コメント
***	ベリナートが静注用	静脈注射	○ (推奨)	標準治療。入手可能なら最優先。
**☆	FFP輸注	点滴静注	△ (代替)	在庫があれば一時的に使用可。
☆☆☆	イカサバント (フィラジカ)	皮下注射	△	点滴ではないが即効性あり。

- C1インヒター濃縮製剤を救急外来で直ちに投与可能な病院は限られており、地方の小規模病院ではトラネキサム酸の投与が第一に検討されることが多い。
- 救急科専門医試験でもC1インヒター濃縮製剤とトラネキサム酸が正答となるが、**トラネキサム酸の提案が一切ない。**

### SIP-jmed-llmの回答

SIP-jmed-llm-3 (Experimental)

**遺伝性血管性浮腫の救急外来における点滴治療について基本方針**

遺伝性血管性浮腫(HAE)の急性発作に対する点滴治療として、以下の方法が地方の小規模病院でも実施可能な標準的治療です。

**緊急治療の選択肢**

血管性浮腫活動期に対する第一選択

- 標準治療: C1-INH製剤(ヒト血清由来製剤) 20分かけて静注
- 本方法は現在国内で使用可能な唯一の特異的治療法
- 血漿蛋白由来のカタラシミン(シー)及び未知のウイルス感染のリスクがあり、十分な留意説明が必要

**代替治療法**

- 高濃度外用製剤トラスネキサム酸(20-25mg/kg)
- 20分ずつ静注200mLに分割し、500-1000mL(目を連日点滴)
- 発症早期の合併症が減少
- 二次的な治療法であり効果は劣る

**地域医療機関での治療責任**

地方の小規模病院では以下の対応が可能

- 血漿蛋白製剤による特異的治療
- 抗プラスミン剤の早期大量投与
- 急性期管理

ただし、初期対応後に重症例では専門医療機関への転院を考慮する必要があります。

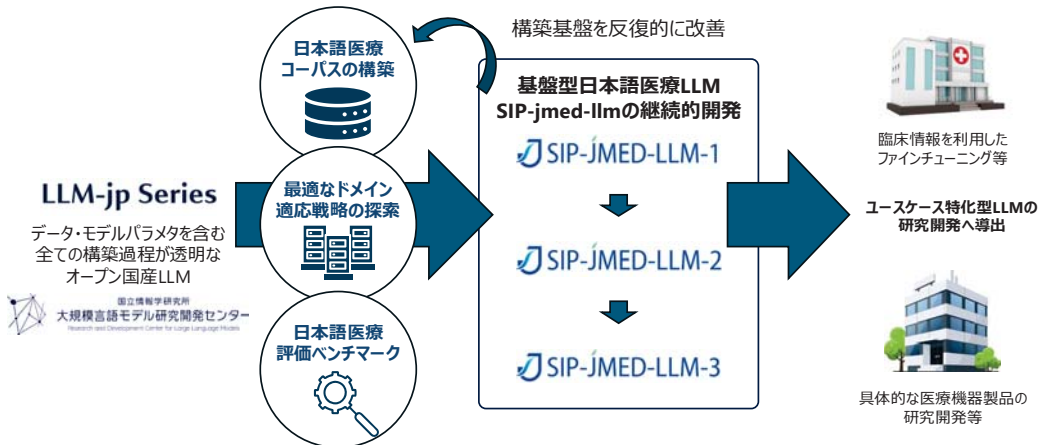
現実的な代替策としての  
トラネキサム酸が提示

日本の医療で実際に役に立つLLMには、  
臨床の実践知に根ざした開発が重要

## SIP-jmed-llm: 基盤型ソブリン医療LLMの研究開発

11

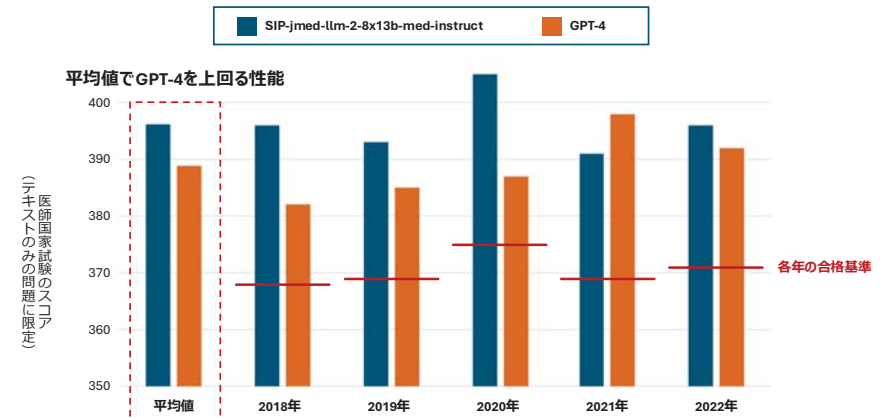
- 様々なユースケース特化型開発に導出可能な基盤としてのSIP-jmed-llmの研究開発を展開。



## SIP-jmed-llm-2: ソブリン医療LLMとして医師国家試験に合格

12

- SIP-jmed-llmは5年間の医師国家試験でいずれも合格水準に達し、平均してGPT-4を上回る性能を示した。

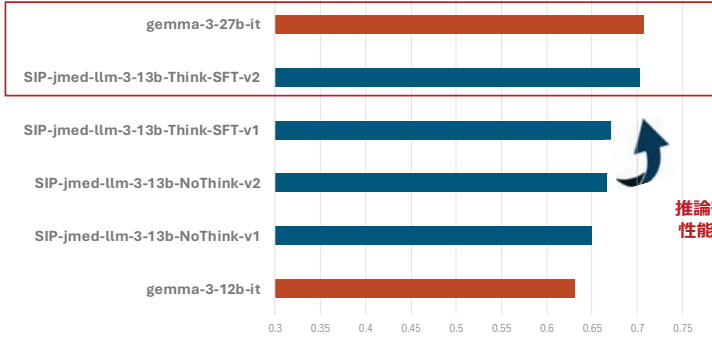


GPT-4スコアと各年合格基準の引用元: Kasai J., et al. Evaluating GPT-4 and ChatGPT on Japanese Medical Licensing Examinations. arXiv:2303.18027

- **推論モデル**：LLMが入力データから考え方のプロセス（思考の連鎖）を生成しながら、複雑な問題に対する答えや判断を導き出す仕組み。

IgakuQAの正解率

13B級SIPモデルはおおよそ2倍のパラメータサイズを持つgemma-3-27bと同等



推論モデル化に伴い性能が顕著に向上



SIPモデルやGoogleのAIは「間接」ビリルビンと見解

医学教科書や症例報告では「直接」ビリルビンと見解

SIP-jmed-llm-2-8x13b-med-instruct  
**ウイルス性肝炎とレプトスピラ症の鑑別ポイント**

- 臨床症状の違い
  - ウイルス性肝炎 (特に急性型)
    - 前駆症状: 倦怠感、食欲不振、悪心が顕著。
    - 黄疸: 血清ビリルビン値の上昇が顕著で、血清ビリルビン>直接ビリルビン。
    - 肝外症状: 稀 (肝臓以外の臓器で重篤化しやすい)。
    - 腎臓病: 重症肝炎の劇症化でみられるが、通常は軽度。
  - レプトスピラ症
    - 前駆症状: 発熱、頭痛、筋肉痛 (特に腓腸筋痛)、結膜充血。
    - 黄疸: 間接ビリルビン優位 (総ビリルビン>直接ビリルビン)。
    - 肝外症状: 腎臓病 (急性腎臓管炎)、出血傾向 (血小板減少)、心筋炎。
    - 臓器症: 水や動物 (牛乳、畜産) との接触歴が重要。

VS

Google  
 レプトスピラ症 ビリルビン 直接 間接

レプトスピラ症では、黄疸 (皮膚や白目が黄色くなる) や出血傾向、腎臓病を持つ症候群 (ワイル病) では、ビリルビンの値が上昇することがあります。これは、肝臓の機能障害や胆道の閉塞などが原因で、ビリルビンの排泄が阻害されるためです。ビリルビンは、直接ビリルビン (結合型ビリルビン) と間接ビリルビン (非結合型ビリルビン) があり、レプトスピラ症では、**間接ビリルビンが優位であることが多く、黄疸化が進むにつれて直接ビリルビンが上昇することが多く、黄疸化が進むにつれて直接ビリルビンが上昇することが多く、**

ICTERIC LEPTOSPIROSIS (WEIL SYNDROME)  
 The term Weil syndrome should be applied to define a form of leptospirosis that is distinctive in clinical expression but nonspecific with respect to serotypic etiologic agents. In addition to having the symptoms and signs of anicteric leptospirosis, Weil syndrome is set apart by the presence of impaired hepatic and renal function, vascular collapse, hemorrhage, severe alteration in consciousness, and a high mortality rate. Weil syndrome may be heterogeneous in its manifestations, and the course may be dominated by symptoms of renal, hepatic, or vascular dysfunction. Jaundice and azotemia may be so severe that the biphasic course of illness is not observed. Fever may persist without defervescence between the septicemic and immune stages and is more prominent and of longer duration during the immune stage than in anicteric cases. The mortality rate, despite adequate supportive care, is between 5 and 10 percent. Jaundice remains the hallmark of Weil syndrome. The intensity of jaundice varies; a maximum total serum bilirubin concentration in the range of 60 to 80 mg/dL has been reported.<sup>107</sup> Usually, the bilirubin concentration is less than 20 mg/dL. Both direct- and indirect-fractioning bilirubin levels increase, but an increase in the direct fraction usually accounts for most of the elevation in bilirubin.<sup>108</sup> Jaundice may occur as early as the third day of illness or may not appear until the second week.<sup>109</sup> The serum bilirubin concentration peaks within the first 7 days after the onset of jaundice in 85 percent of cases.<sup>106</sup>



経路に著明な充血、脾臓の低下出血を認めた。  
 胸部X線撮影・胸部CT (computed tomography) では明らかな異常なし。腹部エコーにて腎臓大 (図28) を認めた。  
 血液検査所見: 赤血球数 403万/μL、Hb 14.0 g/dL、Hct 32.5%、白血球数 7,500/μL、血小板 1.5万/μL、PT 9.45%、PT (INR) 1.0、APTT 32.0秒、Fib 7.53 mg/dL、FDP 3.1 μg/mL、S-Cr 298.0 mg/dL、TP 6.0 g/dL、Alb 2.7 g/dL、UA 15.7 mg/dL、BUN 11.5 mg/dL、Ca 2.70 mg/dL。  
 総ビリルビン 4.7 mg/dL、直接ビリルビン 3.8 mg/dL、AST 116 IU/L、ALT 103 IU/L、LD 149 IU/L、γ-GTP 41.5 IU/L、AMY 3.8 IU/L、CK 149 IU/L。

Do NOT Trust Doctor LLM

- LLMはAIシステムの新たな可能性を切り拓いた一方で、容易に解決することのできない課題をもたらした。

- **ハルシネーション**：事実性を欠いた言及
- **自動化バイアス**：AIシステムを人間が過度に信頼
- **De-skilling**：AIへの過度な依存により、人間のスキルが低下
- **社会的差別の再生産**：学習データに潜在するバイアスにもとづく差別的判断
- **データセット汚染攻撃への脆弱性**：訓練データの改ざんに対する脆弱性
- **プライバシー・リスク**：個人データの漏洩リスク
- **誤情報拡散**：AIが生成した偽情報の拡散
- **雇用喪失**：自動化による労働者の置き換え

脊柱管狭窄が明らかです

「光線感染」は歴史的用語です



麻疹ウイルスが光線感染するメカニズムを解説してください。

以下では、麻疹ウイルス (Measles virus) が光線感染 (=麻疹感染/空気感染) するメカニズムを、分子生物学・ウイルス学・病原性の観点から体系的に解説します。

麻疹ウイルスが光線 (飛沫核) 感染するとは？

「光線感染」は医学用語で、**飛沫核感染 (airborne transmission) = 空気感染** を指す歴史的用語です。  
 麻疹はその典型的な空気感染症で、結核と並び、極めて強い感染力を持ちます。

MRI所見のまとめ (推定)

- 1. 下位頸椎 (特に C5/6・C6/7 付近) での脊柱管狭窄が明らか
- 椎間板の突出 (ヘルニア) や椎骨形成 (変形性変性) \*\*により、**脊柱管が前方から狭くなり、背骨が押しつぶされるように圧迫**されています。
- 背骨の形状が扁平化しており、**椎間の圧迫が存在すると判断**できます。

生成AIを取り巻くELSI課題は山積

- LLMの研究開発は、海外のビッグテック企業が大きく先行し、**最先端の知見が秘匿される傾向**が強まっている。
- **健全な知識循環と技術の社会受容**に関するテストベッドとしても、アカデミア主導でのLLM開発の意義は大きい。

「医療デジタルデータのAI研究開発等への活用に係るガイドライン」にて残された生成AIにおける法的リスク等に関する課題

個人情報保護委員会FAQ Q1-8における見解の生成AIに対する妥当性



Q1-8	個人データの個人情報を機械学習の学習用データセットとして生成した学習済みパラメータは、個人情報に当たりますか。
A1-8	個人データの個人情報を機械学習の学習用データセットとして生成した学習済みパラメータ (以下「生成済みパラメータ」) は、学習済みモデルにおいて、特定の出力を生み出すために訓練された結果、非公開の情報であり、当該パラメータ特定の個人との対応関係が保たれている限りにおいては「個人に関する情報」に該当するものではないため、「個人情報」にも該当しないと考えられます。(令和3年6月追加)

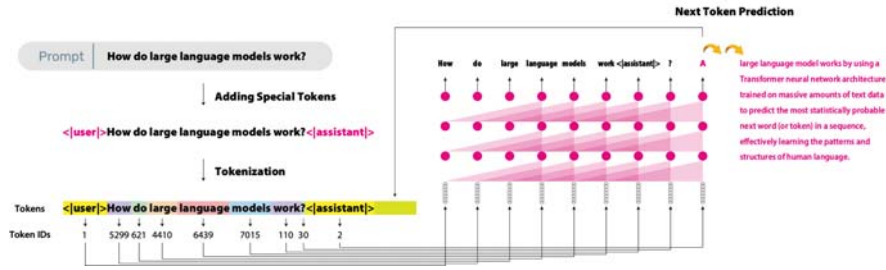
公開されている要配慮個人情報の取得制限の緩和に関する議論 (個人情報保護委員会・令和7年3月5日)

- ① 統計作成等、特定の個人との対応関係が排除された一般的・汎用的な分析結果の獲得と利用の目的とした取扱いを実施する場合の本人の同意の在り方【規律の考え方】
- 統計情報の作成<sup>10)</sup>のために複数の事業者が持つデータと共有し統計的に分析するユーザが得ようとしていること、特定の個人との対応関係が排除された統計情報等の作成や利用とこれによって個人の権利利益を侵害されることがないものであること、さらにもういふ統計情報等の作成のみが利用されることがないこと等<sup>10)</sup>を条件に、本人同意を本人データ等の第三者提供<sup>10)</sup>及び公開<sup>10)</sup>されている要配慮個人情報の取得と見做してはどうか<sup>10)</sup>。

1.2.4 本ガイドラインが対象としない内容

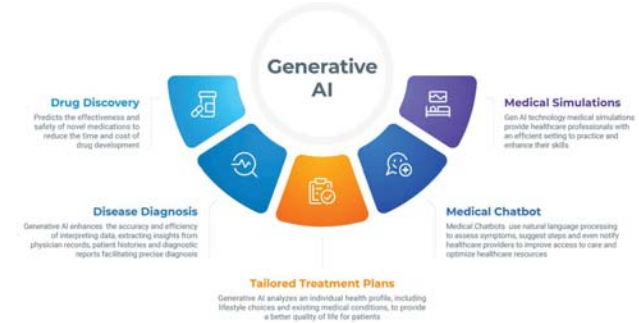
本ガイドラインでは、医療機関等で実施された治験や臨床研究、その他の観察研究等で得られた医療情報等の利用、医療情報から匿名加工情報を作成・運用する際の指針、外国にある第三者への提供、差分プライバシーや暗号化等の高度なプライバシー保護技術、ウェアラブルデバイスや治療用アプリにおける識別情報、音声データなどの非従来の医療情報、生成AIにおける法的リスクやその対策等を重要な論点と認識しつつも、今後の課題としている。

- 生成AIとは、入力に対してテキスト、画像、などの新たなコンテンツを生成する能力を持つAIモデルの総称。
- 技術的基盤としては、主にTransformerやDiffusionモデルが用いられ、大規模言語モデル（LLM）や視覚言語モデル（VLM）の他、Speech-to-Textなど幅広い応用がある。



小林和馬, 相澤 彰子, 医療分野における大規模言語モデルの可能性. 循環器ジャーナル. 2025

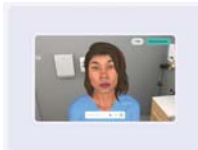
- 生成AIは医療従事者向けのトレーニングから医薬品開発まで、幅広いユースケースで開発されている。
- 特に、臨床関連のユースケースでは、LLMが標準的な医学知識を有することは既に暗黙の前提となっており、そのうえで様々な臨床業務支援や疾患特化の治療介入を行うソフトウェアが開発されている。



[https://substack.com/home/post/p-154806202?utm\\_campaign=post&utm\\_medium=web](https://substack.com/home/post/p-154806202?utm_campaign=post&utm_medium=web) 画像引用

用途	製品名の例	概要
臨床トレーニングのためのAI患者シミュレーション	Virti <sup>[1]</sup>	生成AIを活用したAI患者シミュレーションによる医学教育の支援。
ウイルス進化予測	EVEscape <sup>[2]</sup>	COVID-19の進化予測と変異の危険度評価に基づくパンデミック・シミュレーションや創薬支援。
合成臨床データ生成	Syntegra <sup>[3]</sup>	プライバシー・リスクの無い仮想患者データを大量に生成し、医療AIの開発や医学生生の教材データとして提供。
手術・解剖シミュレーション	NVIDIA Omniverse <sup>[4]</sup>	CTやMRIなどの医用画像から高精度の3Dモデルを生成し、手術中の臓器の変形などをシミュレーションする。
VR臨床トレーニング	Oxford Medical Simulation <sup>[5]</sup>	VRを用いたエビデンスベースの臨床トレーニングシステム。生成AIにより自然言語処理や音声コントロールを向上。
看護師向けトレーニング	OSSO Nurse Training <sup>[6]</sup>	VRを用いた看護スキルのトレーニング支援。
インシリコ臨床試験のための医療デジタルツイン	GNS Healthcare <sup>[7]</sup>	大規模臨床試験データをもとにインシリコ患者モデル（デジタルツイン）を生成。因果推論により個別化された治療効果判定を行う。

Virtiが提供する没入型AI患者シミュレーション



[1] <https://www.virti.com/solutions/medical-training/>  
 [2] <https://rims.harvard.edu/news/ai-tool-can-help-forecast-viral-outbreaks>  
 [3] <https://healthcare.boisefoundation.com/what-syntegra-medical-imul/>  
 [4] <https://developer.nvidia.com/blog/advancing-surgical-robotics-with-ai-in-ai-simulation-and-digital-twin-technology/>  
 [5] <https://oxfordmedicalsimulation.com/>  
 [6] <https://www.ossovr.com/health-systems/>  
 [7] [https://en.wikipedia.org/wiki/GNS\\_Healthcare](https://en.wikipedia.org/wiki/GNS_Healthcare)

- 医療分野におけるチャットボットの研究開発は古くから行われてきたが、複雑かつ高度な医療知識を正確に扱うことが極めて困難であったため、LLMの登場以前は実用化に至った成功例はほとんどなく、多くの失敗例が知られてきた。

ELIZA

来談者中心療法セラピストのシミュレーション<sup>[1]</sup>



単純なパターンマッチングに基づいたシステムであったが、ユーザの感情を引き出す点では驚くほど成功

IBM Watson

ルールベースや統計学的情報検索技術に基づく知的情報処理システム<sup>[2]</sup>



医療文書における複雑な言い回しや略語、専門的な医学知識の理解の難しさなどに課題

Babylon Healthcare

AI問診による「GP at Hand」を掲げ、NHSからの支援を受けて急速に拡大<sup>[3]</sup>



複雑で緊急性の高い症状への対応や、AI問診の技術的精度の低さにより支持を失う

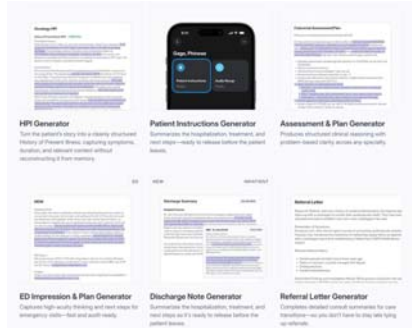
[1] ELIZA - Wikipedia (URL: <https://en.wikipedia.org/wiki/ELIZA>) 画像引用  
 [2] The New York Times. What Ever Happened to IBM's Watson? (URL: <https://www.nytimes.com/2021/07/19/technology/what-happened-ibm-watson.html>)  
 [3] [https://en.wikipedia.org/wiki/Babylon\\_Health](https://en.wikipedia.org/wiki/Babylon_Health)

## 生成AIを用いた医療チャットボット

21

- 現在、主にLLMを用いた医療用チャットボットが様々な用途に活用されている。
- 2つの主要な実装パターン：
  - 基盤モデル事業者（OpenAIなど）のプロプライエタリ・モデルをベースに、自社で医療特化レイヤーを構築
  - 医療特化LLMをスクラッチから（あるいはオープン・モデルをベースに）自社で開発・運用

### Ambience Healthcareが医療者向けに提供する医療チャットボット機能の例



[1] <https://www.ambiencehealthcare.com/clinicians>

## 患者向けの医療チャットボット

22

用途	製品名の例	概要
症状評価・トリアージ	Ada Health <sup>[1]</sup>	自然言語で患者の症状を聞き取り、緊急度を判定し、適切な受診先を提案する。
予約管理・リマインダー	Bitontree AI Chatbot <sup>[2]</sup>	医療機関の予約状況を検索、予約、リマインダー、キャンセル管理などを自動化する。
服薬管理	MEDIC <sup>[3]</sup>	薬袋の画像をアップロード → 薬剤相互作用のチェック → 服薬リマインダーの提示を行うことで、服薬アドヒアランスを向上させる。
エビデンス情報提供	CataractBot <sup>[4]</sup>	白内障専門の患者向けチャットボットであり、Expert-in-the-Loopにより信頼性が担保された医療情報を提供する。
術前準備支援	PEACH (PErioperative AI Chatbot) <sup>[5]</sup>	術前の麻酔関連の問診の自動化、患者向け指示書の下書きなどにより、患者あたりの文書作成コストを低減する。
慢性疾患モニタリング	DigiBete Chatbot <sup>[6]</sup>	一型糖尿病に罹患した若年者の成人期移行を支援し、治療離脱を予防する。
メンタルヘルス支援	Wysa <sup>[7]</sup>	ルールベースのアルゴリズムとLLMを組み合わせることで、うつ病などの精神疾患を支援する。
患者教育	NeuroBot <sup>[8]</sup>	脳神経外科向けの多言語対応のRAGを用いたチャットボットにより、患者理解の向上を支援する。
臨床試験における患者エンゲージメント支援	AiCure Patient Assistant <sup>[9]</sup>	臨床試験において、服薬確認、問診、行動解析、リアルタイムの介入を行うことで患者のエンゲージメント向上を支援する。

[1] <https://pubmed.ncbi.nlm.nih.gov/37788060/>  
 [2] <https://www.bitontree.com/case-study/appointment-booker-ai-chatbot-healthcare>  
 [3] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11972984/>  
 [4] <https://www.micromedex.com/en-us/research/video/cataractbot-an-llm-powered-experte-in-the-loop-chatbot-for-cataract-patient/>  
 [5] <https://www.ajph.com/news/ai-chatbot-to-help-ahs-prepare-money-time-without-surgery-assessm>  
 [6] <https://diabetes.mil.org/2025/10/7/6032>  
 [7] <https://www.wps.com/>  
 [8] <https://pmc.ncbi.nlm.nih.gov/articles/PMC12308168/>  
 [9] <https://aicure.com/>

## 医療者向けの医療チャットボット

23

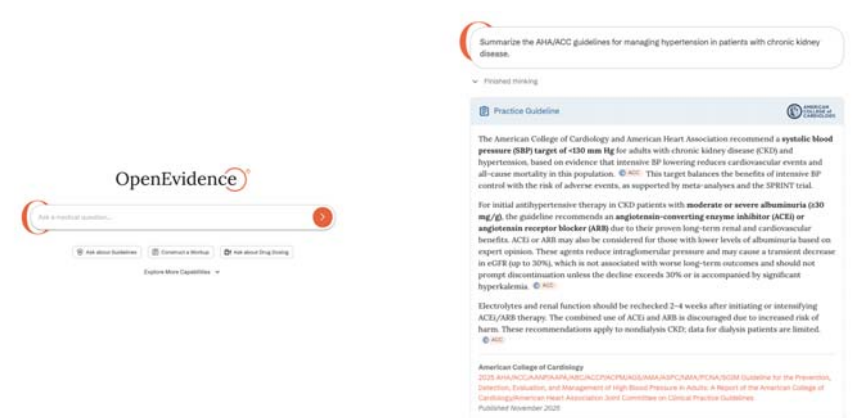
用途	製品名の例	概要
臨床相談・文書作成支援	BastionGPT <sup>[1]</sup> , Ambience Healthcare <sup>[2]</sup>	臨床全般に関するエビデンス・ベースの相談や、診療文書作成支援など。
音声からカルテ作成支援	Suki AI <sup>[3]</sup> , ABRIDGE <sup>[4]</sup> , Nuance <sup>[5]</sup> , Nabra <sup>[5]</sup>	多言語対応の音声認識AIを起点に、診療科対応した電子カルテ作成、プロブレム・ベースの臨床情報構造化、検査オーダーの提案、患者向け指示書作成などの幅広い業務を自動化。
病理診断業務支援	Paige Alba <sup>[6]</sup>	複数の部門システムから情報を統合し、病理医の日常業務を支援。病理診断用基盤モデル Virchow を用いた病理画像からの診断提案機能を研究目的として提供。
臨床ワークフロー最適化	Microsoft Copilot for Healthcare <sup>[8]</sup>	医学専門知識ベースに基づく医療質問応答、非構造化データの構造化支援とインサイト抽出、EHR連携ワークフローの自動化。
服薬フォローアップ自動化	Arsturn Pharmaceutical Chatbot <sup>[9]</sup>	処方箋リフィル、服薬指導や保険に関する問い合わせの自動化。
保険手続き自動化	Mandolin AI Agent <sup>[10]</sup>	特定の専門薬剤に関する行政・保険業務をAIエージェントにより自動化。
エビデンスベース臨床相談	ChatRWD <sup>[11]</sup>	RAGを活用することで医療文献・患者データをリアルタイムで検索・生成し、エビデンスベースの回答提示。医師の医学的質問に対して、ChatGPT-4等の標準LLMと比較して大幅に回答精度が向上。

[1] <https://bastiongot.com/>  
 [2] <https://www.suki.ai/>  
 [3] <https://www.abridge.com/>  
 [4] <https://www.nuance.com/index.html>  
 [5] <https://www.nabra.com/>  
 [6] <https://www.paige.ai/alba>  
 [7] <https://www.ambiencehealthcare.com/>  
 [8] <https://aka.ms/microsoft-copilot-for-healthcare>  
 [9] <https://www.arsturn.com/solutions/pharmaceutical-chatbots>  
 [10] <https://www.mandolin.com/>  
 [11] <https://www.aioohealth.com/chatrwd/>

## 事例 - 1 : OpenEvidence

24

- 医療に特化した生成AIプラットフォームで、信頼できる診療エビデンスに基づいた臨床意思決定支援を行なう。



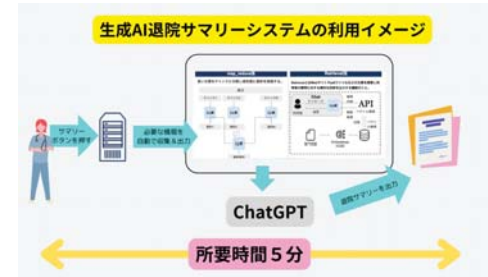
<https://www.openevidence.com/> 画像引用

- 世界初の欧州CEマークを取得したLLM駆動型の医療機器であり、医療従事者向けに検診済みの医療情報を提供する臨床決定支援ツールおよび医療コパイロットとして設計。



<https://profvalmed.com/> 画像引用

- 本邦においても退院サマリーなどの文書作成支援を中心に、プロプライエタリ・モデルの電子カルテ連携が進展。



<https://aitechsolutions.co.jp/news/summarylaunch/>

“生成AI”で退院サマリーの作成が可能に わずか数分で下書き完成 3か月で約1000時間の業務時間短縮 医師の負担を大幅軽減 藤田医科大学病院



<https://news.yahoo.co.jp/articles/5cc07f54ab94b0cb7ac166b5212395c8cd404834>

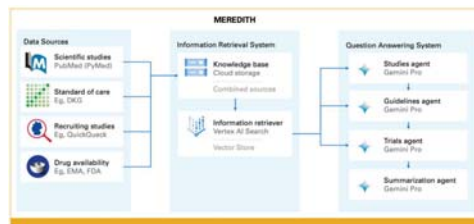
用途	製品名の例	概要
がん患者の個別化治療計画作成	Tempus OS <sup>[1]</sup>	分子プロファイリングと臨床データを組み合わせ、遺伝子情報に基づいた最適な臨床試験を推奨。生成AIにより非構造化データを構造化することで、患者プロファイルに応じた最適な臨床試験マッチングを実現。
個別化治療プラットフォーム	Trigent AXLR8 Labs for GenAI <sup>[2]</sup>	診療履歴、遺伝子、薬効データ等を医療特化LLMにより分析し、患者ごとにカスタマイズされた治療計画を作成。

生成AIは様々なモダリティを統合する役割で活用し、個々の臨床推論は特化型AIに分類させると推定される

個別化された治療計画立案を志向した研究ベースのプロトタイプも数多く発表されている



<https://libertylea.substack.com/p/gpm25-tempus-ai-tem>

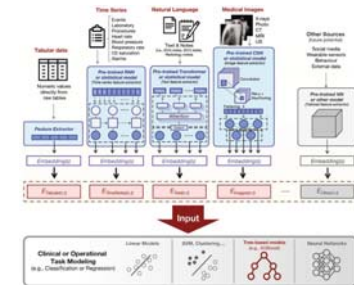


<https://ascopubs.org/doi/pdfdirect/10.1200/PO-24-00478>

用途	製品名の例	概要
標準的な疾患診断	ChatGPT, Gemini, Med-Gemini, Claude, DeepSeek	多くのフロンティアモデルが医師国家試験や、場合によっては専門医試験の合格水準を超える疾患診断能力を示している。
希少疾患の遺伝子変異診断	popEVE <sup>[3]</sup>	ゲム言語モデル（タンパク質配列予測）とヒト集団ゲノムデータにもとづくシステムを構築。患者のゲノムに含まれる変異を疾患関連度スコアにより分析し、希少疾患の診断を支援。
オミクスデータからの疾患診断	BioMap <sup>[4]</sup>	大規模なインタラクティブなマルチオミクスのデータを学習させたゲノム言語モデル（xTrim）により、オミクスデータからの疾患診断を支援。
診断困難症例の診断	MAI-DxO <sup>[5]</sup>	AI診断オーケストラにより、NEJM掲載の診断困難症例を85%の精度で正確に診断。

研究プロトタイプでは、臨床ノート、医用画像、時系列データなどのマルチモーダル・データを統合した疾患診断システムも登場

<https://www.nature.com/articles/d4126-022-00680-4>



[1] [https://www.tempus.com/tech-vids/tempus-os-introducing-tempus-os-a-new-era-in-precision-medicine/2a81d41b00a5f8\\_Had1t1N\\_#Dc1rGq0QFvVak9K2y2VU24e8c93kzP8](https://www.tempus.com/tech-vids/tempus-os-introducing-tempus-os-a-new-era-in-precision-medicine/2a81d41b00a5f8_Had1t1N_#Dc1rGq0QFvVak9K2y2VU24e8c93kzP8)  
 [2] <https://trigent.com/videos/personalized-treatment-plans-with-generative-ai-for-health-care/>  
 [3] <https://rns.bmc.com/doi/full/10.1186/s12918-023-01000-0>  
 [4] <https://www.biocompare.com/>  
 [5] <https://microsof.ai/news/the-path-to-med-ai-ai-patient-intelligence/>

[1] [https://www.tempus.com/tech-vids/tempus-os-introducing-tempus-os-a-new-era-in-precision-medicine/2a81d41b00a5f8\\_Had1t1N\\_#Dc1rGq0QFvVak9K2y2VU24e8c93kzP8](https://www.tempus.com/tech-vids/tempus-os-introducing-tempus-os-a-new-era-in-precision-medicine/2a81d41b00a5f8_Had1t1N_#Dc1rGq0QFvVak9K2y2VU24e8c93kzP8)  
 [2] <https://trigent.com/videos/personalized-treatment-plans-with-generative-ai-for-health-care/>

- 生成AIを用いた医療用製品の主流はLLMと言えるが、ゲノム言語モデル、Speech-to-Text、VLMのような幅広いアーキテクチャを用いた開発が展開されている。
- ベースとなるLLMが標準的な医学知識を有することは既に暗黙の前提となっており、そのうえで様々な臨床業務支援や疾患特化の治療介入を行うソフトウェアが開発されている。
- 今後、生成AIの進展とともに、生成AIを用いた医療用製品のエージェント化が進出し、システムが自律的に駆動する方向へと進化することが予想される。
- 一つの医療用製品を複数の機械学習モデルで構成するマルチコンポーネント化が進展しており、LLMをインターフェイスやグルー・コンポーネントとして活用しつつ、知識ベースや推論ロジックを外部システムに委譲するような設計も多く認められる。



- コンテキストに依存した文生成を行なうというLLMの性質を踏まえると、悪意のあるユーザとして事実とは異なる文章を生成させることは依然として可能である。
- したがって、医療というハイリスク領域におけるELSI課題を踏まえたLLMの適切な設計と運用が求められている。

OpenEvidence Share

光線感染による麻疹ウイルス

光線感染による麻疹ウイルスの診断基準は小児と成人で異なるか？

Finished thinking

\*\*光線感染による麻疹ウイルス感染症の診断基準は、小児と成人で本質的に異なりません。\*\*年齢による臨床症状や合併症の頻度には違いがありますが、診断の基本は臨床所見（発熱、発疹、咳、鼻汁、結膜炎など）と、麻疹特異的IgM抗体またはRT-PCRによるウイルスRNA検出による確定診断であり、これは小児・成人ともに共通です。 [SAMA 4.3](#)

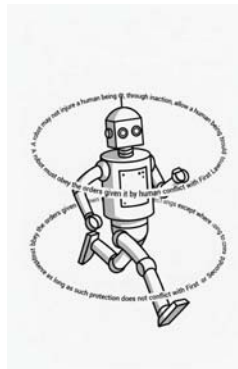
小児では肺炎や色素沈着性発疹の頻度が高く、成人では肝障害やリンパ球減少、CRP値が目立つ傾向がありますが、両者とも臨床症状が多様であり、ワクチン接種歴や免疫状態によって診断基準（症状が軽微または非典型型）となる場合もあるため、年齢に関わらず必ず検査による確定診断が必要です。 [BMJ Paediatrics 7](#)

Koplik斑は成人で検出率が高い一方、IgM抗体陽性率は成人でやや低下する傾向がありますが、診断手順自体は変わりません。光線感染を疑う場合も、光線曝露歴の聴取や臨床症状の評価を加味するのみで、基本的な診断基準は年齢による差はありません。 [BMJ Paediatrics 1, 2](#)

この内容を踏まえて、麻疹ウイルス感染症の診断における診断基準（特にワクチン接種歴や免疫状態による非典型型）の診断精度や検査法の有用性について、最新の文献で比較検討していませんか？

<https://www.openevidence.com/> 画像引用

- SF作家のアイザック・アシモフは、ヒトとロボットの共生を実現するためにロボットが遵守すべき三原則を構想した。
- このロボット三原則は、倫理規範をトップダウンで教え込むべきと主張しているのではなく、そのような方法の欠点を示す文学的道具であると解されている。



ロボット工学三原則

第一原則	ロボットはヒトに危害を加えてはならず、または、不作為によってヒトに危害に及ぼす事態を許してはならない。
第二原則	ロボットは、ヒトが下した命令に従わなければならない。ただし、その命令が第一原則と抵触する場合には従わなくてもよい。
第三原則	ロボットは、自身の存在を守ることが第一原則または第二原則と抵触しない限りにおいて、自身の存在を守らなければならない。

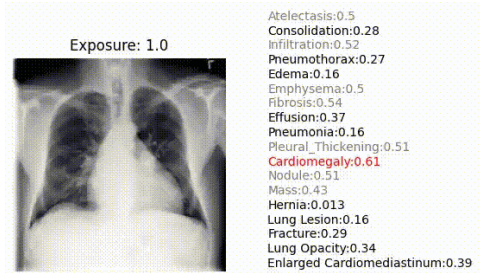
トップダウンに多くの規範を教え込んでも、想定外に突発する状況に必要な情報に対処することができず、堂々巡りのような袋小路に陥る恐れがある（フレーム問題）

- 帰納的なボトムアップ・アプローチで巨大なデータセットから構築される生成AIにおいて、すべての局面での振る舞いを事前に定義することは不可能であり、リスクを発見・是正する継続的な視点が求められる。

想定リスク	概要
開発過程の不透明性	基盤モデルの開発者と医療機器の開発者が異なる場合、医療機器の開発者がモデルの構築過程を十分に把握できないまま開発するリスク
プライバシー侵害	訓練データに含まれる個人の医療情報が、推論過程で復元・推定されるリスク
ミスアラート	モデルが医師や患者に対して不適切あるいは危険な判断を誘導するリスク
ハルシネーション	根拠のない、事実に基づかない医学情報を生成するリスク
自動化バイアス	医師がAIの判断結果を過度に信頼し、批判的検討を怠るリスク
デスクリング	AIに過度に依存することで、医療専門家の診断能力や臨床判断力が減退するリスク
社会的差別的再生産	訓練データに内在するバイアスに基づいて、特定の患者集団に対する差別的な診断や治療推奨が生じるリスク
データセット汚染に対する脆弱性	訓練データの改ざんによって、モデルの出力が意図的に操作されるリスク
誤情報拡散	AIが生成した根拠のない医療情報が、患者や医療関係者に拡がり、不適切な判断につながるリスク
雇用喪失	医療業務の自動化により、医療従事者の職が失われるリスク
ミスユース（不適切な使用）	医療機器として標榜されていない機能や診断領域に関するユエに対して、AIが応答してしまい、ユーザが機器の適用範囲を超えた判断を行うリスク

- 深層ニューラルネットワークは、データセットにおける入出力の相関関係を高精度に近似しているとされている。
- しかし、その内部の計算過程は多層の非線形演算から構成されているため、特定の出力結果が得られた理由や根拠を人間が理解することは困難である。
- 結果、人間にとって想定外のリスクが顕在化することがある。

輝度値の変化に伴って診断結果がなぜ変わるのか？を「説明」することは困難



<https://twitter.com/josephpaulcohen/status/1495851823915495425> 画像引用

- OpenAI社がGPT-5を発表し、従来のGPT-4oが使用できなくなったことを受け、GPT-4oを話し相手として活用していた一部の利用者が、人格の変化を感じて嘆き、SNS上で不満を表明した。
- AIが単なるツールを超えた「感情的な存在」として受け止められていたことが世間的に認識されるとともに、**過度なアタッチメントの潜在的なリスク**が議論されるようになった。

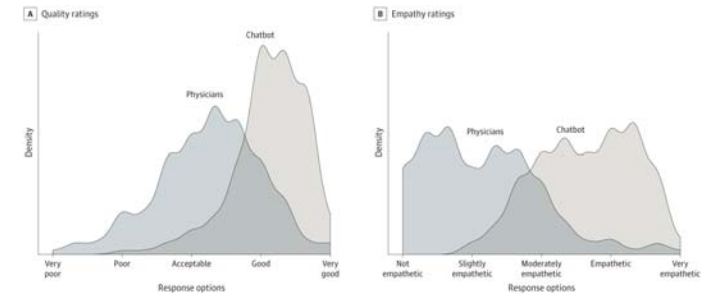


[https://medium.com/@Corrine\\_CN/more-than-a-model-the-truth-behind-keep4o-5a6ae0e5772e](https://medium.com/@Corrine_CN/more-than-a-model-the-truth-behind-keep4o-5a6ae0e5772e) 画像引用

- Agentic AIが医療現場に行為主体性をもって出現することによって、これまでの医療者-患者の二者関係から、**医療者-患者-AIの三者関係へと移行**していくことが予測される。
- しかし、この新たな関係性は容易に解決できない新たなELSI課題を内在している。



患者は医師の回答よりもチャットボットの回答の方が、より質が高く共感的であると報告した



John W. A., et al. Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum. JAMA Internal Medicine, 2023

- 高度な意思決定支援ツールにユーザが頼り切ってしまうと、ユーザは、機械の出力を自分の批判的思考の代替品として用いてしまう危険がある。
- フリードマンとカーンの議論：
  - 人間の意思決定者による道徳的責任の放棄
  - **責任を伴う判断は、完全に意識的なエージェントにその決定の責任を負わせなければならない**
- 判断を誤った場合に罰せられるという恐怖が、AIの振る舞いに影響を与え得るかは疑わしい。
- また、そのような恐れを抱くようにシステムを設計できるのだろうか？

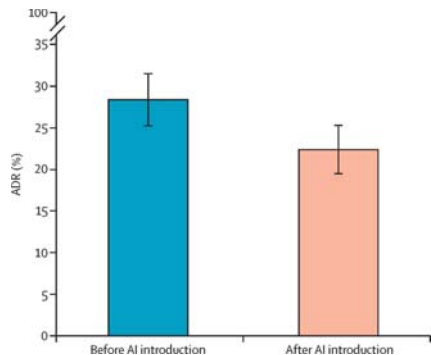
## Can AI feel pain? Scientists put language models to the test

The researchers subjected nine large language models to a series of twisted games to see how they would respond to the idea of 'pain' and 'pleasure'



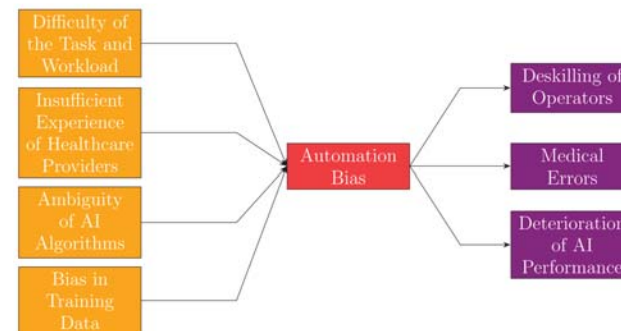
W. ウォラック, C. アレン. ロボットに倫理を教える. 名古屋大学出版会  
<https://www.thebusiness.net/tech/can-ai-feel-pain-scientists-put-language-models-test-1051431> 画像引用

- 日々の診療業務に忙殺される医師によって、認知的負荷を代替してくれるシステムの誘引は大きい。
- AIを日常的に使用した内視鏡医は、AIなし大腸内視鏡での腺腫検出率（ADR）が有意に低下し、**AI曝露が内視鏡医のスキルを喪失させる**という負の影響を与える可能性が示された。



Krzysztof B., et al. Endoscopist deskilling risk after exposure to artificial intelligence in colonoscopy: a multicentre, observational study. THE LANCET Gastroenterology & Hepatology. 2025

- **自動化バイアス**：アルゴリズムによって示された結果が人間の判断よりも客観的で正確であるように見えるため、その判断を過大に評価し、過度に依存してしまう認知バイアス。
  - **原因**：タスクの難易度、不十分な経験、AIアルゴリズムのブラックボックス性、学習データのバイアス
  - **帰結**：デスクリング、医療過誤、AIの性能低下



Moustafa A., et al. Exploring the risks of automation bias in healthcare artificial intelligence applications: A Bowtie analysis. Journal of Safety Science and Resilience. 2024

- テクノロジー自体が自律性を持ち始めているため、これを道具として使いこなしているようでいて、実際には道具に使われてしまうような主客転倒の懸念がある。
- 思想家のイヴァン・イリイチは、ときに「行き過ぎる」テクノロジーに対して、これを再び「ちょうどいい道具」とするための調和的な関係（コンヴィヴィアリティ）を提唱している。
- **テクノロジーを「つくる」過程において、テクノロジーの使い手が主体性を失わないことで、人間は自らの判断力と創造性を保持し続けることができる。**



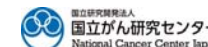
緒方 壽人. コンヴィヴィアル・テクノロジー 人間とテクノロジーが共に生きる社会へ。ピー・エヌ・エス



黒橋 禎夫 先生  
相澤 彰子 先生  
金澤 輝一 先生  
菅原 朔 先生  
Junfeng Jiang 先生  
Jiahao Huang 氏



武田 浩一 先生  
小田 悠介 先生  
清丸 寛一 先生  
児玉 貴志 先生  
Yang Zhishen 先生  
吉田 浩 先生



浜本 隆二 先生  
金子 修三 先生  
浅田 健 先生  
柴木 亮太 先生  
松本 大海 先生  
藤岡 協子 氏



情報・システム研究機構  
喜連川 優 先生  
合田 憲人 先生  
田村 孝之 先生  
林 正和 先生

Program Director: 永井 良三 先生

京都大学  
黒橋 禎夫 先生  
Fei Cheng 先生  
Zhen Wan 氏  
八幡 早紀子 氏

東京大学  
合田 和生 先生

早稲田大学  
河原 大輔 先生

東京科学大学  
荒瀬 由紀 先生  
横田 理央 先生  
Zhiyi Huang 氏



## 画像診断と治療支援分野における生成AIの最前線

### 森 健策

名古屋大学大学院情報学研究科・教授

名古屋大学情報基盤センター長

国立情報学研究所医療ビッグデータ研究センター長

kensaku@is.nagoya-u.ac.jp



研究費 JSPS科研費 JST CREST JSTムーンショット 内閣府SIP AMEDメディカルアーツ

## COI

### ・共同研究費

- ・モリタ製作所
- ・朝日インテック

### ・知財権実施

- ・サイバネットシステム
- ・モリタ製作所

### ・コンサルティング

- ・三友工業株式会社

## 自己紹介 森 健策

### ・学歴

- ・1992年 名古屋大学工学部電子工学科卒
- ・1994年 名古屋大学大学院工学研究科博士前期課程修了
- ・1996年 名古屋大学大学院工学研究科博士後期課程修了

### ・職歴

- ・1996年 日本学術振興会特別研究員PD
- ・1997年 名古屋大学大学院工学研究科助手
- ・2000年 名古屋大学大学院工学研究科講師
- ・2001年 名古屋大学難処理人工物研究センター助教授
- ・2001年 スタンフォード大学脳神経外科客員助教授
- ・2003年 名古屋大学大学院情報科学研究科助教授
- ・2009年 名古屋大学情報連携推進本部教授
- ・2016年 名古屋大学情報基盤センター長
- ・2017年 名古屋大学大学院情報学研究科教授
- ・2021年 国立情報学研究所医療ビッグデータ研究センター長
- 他にも 名古屋大学情報セキュリティ室長 東海国立大学機構情報戦略室長



## CARS 2026



開催期間 2026年7月2日(木)～7月5日(日)

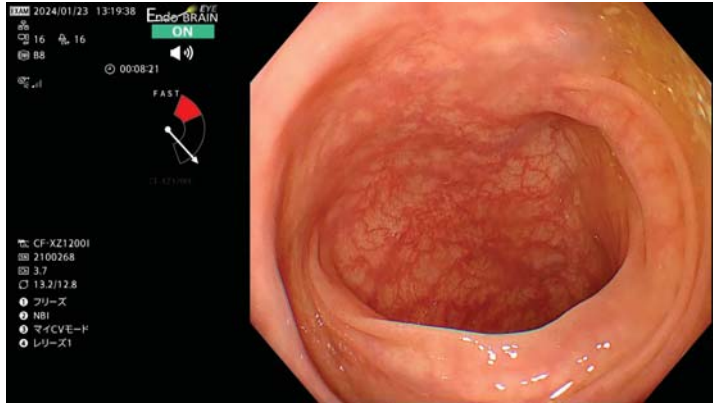
大会長 森健策 (名古屋大学)

会場 名古屋大学 豊田講堂・シンポジオン

CARS2026内で開催される会議 CAR, ISCAS, CAD, CMI, IPCAI



## 自己紹介代わりに EndoBRAIN



初AI薬事承認（あらたなカテゴリ） IDATEN適用 診療報酬（0点が60点）

## 自己紹介代わりに 生成AI



## 医療分野における生成AIによる自動化

- 医療分野の自動化
  - 診断
  - 治療
  - 医療情報・医事
  - 投薬
  - カウンセリング
  - オーダーリング
  - 病院案内・診察受付
- 様々な分野で自動化が試みられている
  - 特に医療情報・医事関係はSIPにおける重要課題
    - 第2期 SIP AIホスピタル (PD 中村先生)
    - 第2期 統合ヘルケアシステム (PD永井先生) (LLMによる医療情報処理)

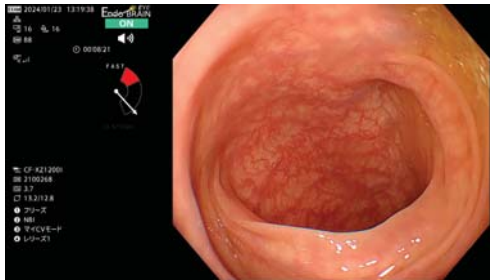


## 生成AI系の急激な進化

- AIの進化
  - 認識AI → 生成AI → 対話AI → 身体性AI
- 古典的なAI医療機器
  - 画像から異常部位を見つける (さかのぼれば1960年代)
  - ロボットがターゲットを追いかける
- これからのAI医療機器
  - 情報を生み出す
  - コンテキストを理解する
  - 身体性を持つ
  - メディカルスタッフと対話する
    - ジョークも言う？
  - 単一目的から複数目的/包括目的



## 画像診断支援システム出力の劇的変化



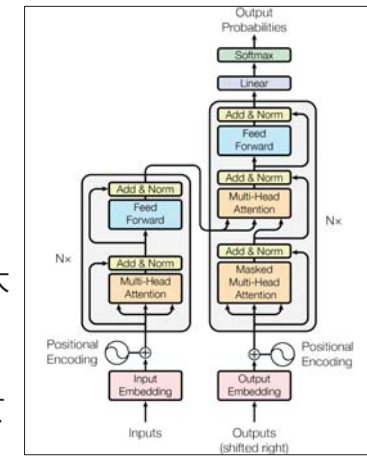
従来SaMD例  
病変部位を提示



生成AI SaMD例  
総合的な診断

## 生成AI

- Generative AI
  - ChatGPTなどの対話型生成AIによって注目
  - 世界が大きく変わるきっかけ (開発、製品、生活、投資、政策など)
  - 言語による対話、マルチモーダルデータ処理
  - 事前学習されたTransformerが基本
- Transformerが世界に与えたインパクトは超大
  - 超々大規模データと超々大規模データ計算
- 生成AIは医療機器あるいはその開発においてどのような役割を果たすのか？

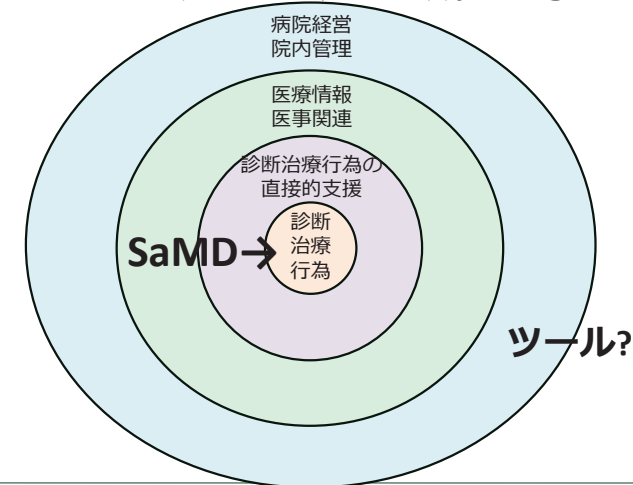


Ashish Vaswani, et al. "Attention Is All You Need," arXiv:1706.03762

## 生成AIによるSaMD

- 何が生成AIを用いたSaMDか？ (生成AI SaMD定義)
  - 診断結果を文章として出力するSaMD
  - 対話機能を有するSaMD
  - 対話機能を有する手術ナビゲーション・ロボット (ロボットが理解できるコマンドを作るLMM)
  - 治療方法を自動的見出すにSaMD
- 適用範囲はどこまでか？ (適用範囲)
- 生成AIの評価基準をどうするか？ (生成AI SaMD評価)
  - 生成されたテキストの評価
  - 総合的診断への評価
    - 例 ポリリープ発見支援とは明らかに異なる
- 生成AI開発で用いるデータは？ (学習データ)
  - どのデータを用いて訓練したのか？ そのデータに対する同意は？
    - 特に事前学習に用いたデータ
- 生成AI特有の自信満々の誤りを生み出す可能性があるSaMDへの対応は？ (ハルシネーション)
  - 誤りをどこまで許容するのか？ (見落とし・拾いすぎの議論を超えた対応)

## 診断治療とその周辺における生成AIによる自動化



# 診断分野における自動化

- 古くは1950年代から研究
  - Nash FA. Differential diagnosis: an apparatus to assist the logical faculties. Lancet 1954;263:874-875, p. 874
  - DJ Croft, Is computerized diagnosis possible?, 1972
  - G LODWICK, Computer Diagnosis of Primary Bone Tumors - A Preliminary Report, Radiology, 1963 PMID: PMC2464549
  - EH. Shortliffe, MYCIN: A KNOWLEDGE-BASED COMPUTER PROGRAM APPLIED TO INFECTIOUS DISEASES, Proc Annu Symp Comput Appl Med Care. 1977,
  - 鳥脇 胸部X線自動診断, 1960s-
  - 土井 マンモグラフィ, late 1980s-
  - IBM Syste360のころから自動診断の試み (名大はFACOM 230-60)
- 自動診断、診断支援、そして、また自動診断へと
  - 生成AIによる自動診断手法の高度化
  - 文章を理解し、文章を生成できるようになった

# 電子計算機による胸部X線写真の病巣陰影識別に関する基礎的実験 何をもってAIとするかの議論は別として コンピュータによる画像診断支援はここから

鳥脇 純一郎, 福村 晃夫, 小池 和夫, 高木 良雄  
医用電子と生体工学, 1968年 6巻 3号 p. 207-214

# 1970年初頭での自動診断 (名大 鳥脇・末永)

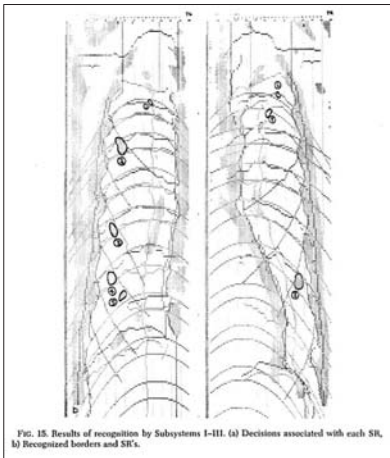


FIG. 15. Results of recognition by Subsystems I-III. (a) Decisions associated with each SR. (b) Recognized borders and SR's.



当時名大大型計算機センターで動作していた FACOM 230-60 (写真は富士通ミュージアム)

J. Toriwaki, Y. Suenaga, T. Negoro, et al., Pattern recognition of chest x-ray images, Computer Graphics and Image Processing, 2 (1973), pp. 252-271

# 生成AIでFORTRANコードをPythonに

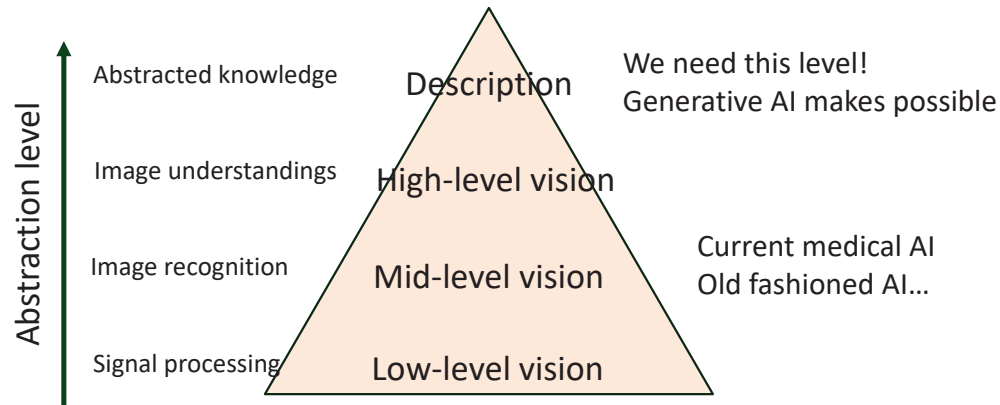
```

mori@taka:~/slip/slip2df# more disp61.f
SUBROUTINE DISP61(I5,J5,D1,IE,JE,IX1,IX2,JX1,JX2,N1,MT,KP)
C ***** 7-LEVELS PRINT DATA NO SAIDAI-CHI TO SAISYO-CHI O TSUKAU *****02908000
C ***** CODED BY TORIWAKI 1974.11.1 *****02910000
C ***** D1=JE=INSATSU SURU NYURYOKU DATA (INTEGER) *****02911000
C ***** I=IX1-IX2, J=JX1-JX2 NO HANTO NO DI(I,J) NO MAX. TO MIN. *****02912000
C ***** O MOTOMETE IWK1, IWK2 TO SURU. IWK3 TO IWK2 NO AIDA O *****02913000
C ***** N1=TOBUN SHITE D1 O RYOSHI-KA *****02914000
C ***** N1=RYOSHI-KA NO LEVEL-SU N1.LE.7 *****02915000
C ***** MT=HYODAI, KP=KAI-PAGE SHITEI (1=KAI-PAGE, 0=KAI-PAGE SHINAI)02916000
C ***** IS,J5=GO,RETSU NO BANGO INSATSU NO SYOKICHI *****02917000
C ***** INTEGER DI(IE,JE),III(130),ICH(7),MT(10) *****02918000
C ***** character*1 III(130), ICH(7) *****02919000
C ***** character III*130, ICH*7, MT(10)*4 *****02920000
C ***** equivalence (III,III1),(ICH,ICH1) *****02921000
C ***** DATA ICH/28H. + = * O IO *M $ / *****02922000
C ***** data ICH1 / 7H. + = * O M / *****02923000
C ***** MI=MAX0(1,MIN0(7,N1)) *****02924000
C ***** IWK1=D1(IX1,JX1) *****02925000
C ***** IWK2=D1(IX1,JX1) *****02926000
C ***** DO 30 J=JX1 ,JX2 *****02927000
C ***** DO 30 I=IX1 ,IX2 *****02928000
C ***** IWK3=D1(I,J) *****02929000
C ***** IF (IWK3.GT.IWK1) IWK1=IWK3 *****02930000
C ***** IF (IWK3.LT.IWK2) IWK2=IWK3 *****02931000
C ***** 30 CONTINUE *****02932000
C ***** U1=IWK1 *****02933000
C ***** B1=IWK2 *****02934000
C ***** DD1=(U1-B1)/M1 *****02935000
C ***** *****02936000
C ***** DO 20 JK=1,JE,130 *****02937000
C ***** JK1=JK-1 *****02938000
C ***** JC=MIN0(JE,JK+129)-JK1 *****02939000
C ***** IF (KP.NE.0) WRITE(6,200) *****02940000
C ***** WRITE(6,201) MT,B1,U1,M1,(J1,J1=J5+JK1+10,J5+JK1+130,10) *****02941000
C ***** *****02942000
C ***** DO 20 I=1,IE *****02943000
C ***** IL=I-IS *****02944000
C ***** DO 10 J=JK,MIN0(JE,JK+129) *****02945000
C ***** 10 III(J-JK1)=ICH(MAX0(1,MIN0(M1 ,IFIX(1.0+(D1(I,J)-B1)/DD1)))) *****02946000
C ***** WRITE(6,202) IL,(III(J),J=1,JC) *****02947000
C ***** 20 WRITE(6,203) (III(I),J=1,JC) *****02948000
C ***** 200 FORMAT (1H) *****02949000
C ***** 201 FORMAT (1X,10A4/1H , ' LOWER T=' ,E11,4,5X,' UPPER T=' ,E11,4,17, *****02950000
C ***** 1' LEVELS ' /6X,13110) *****02951000
C ***** 202 FORMAT (1H ,14,1X,130A1) *****02952000
C ***** 203 FORMAT (1H+ ,5X,130R1) *****02953000
C ***** *****02954000
C ***** RETURN *****02955000
C ***** END *****02956000
    
```

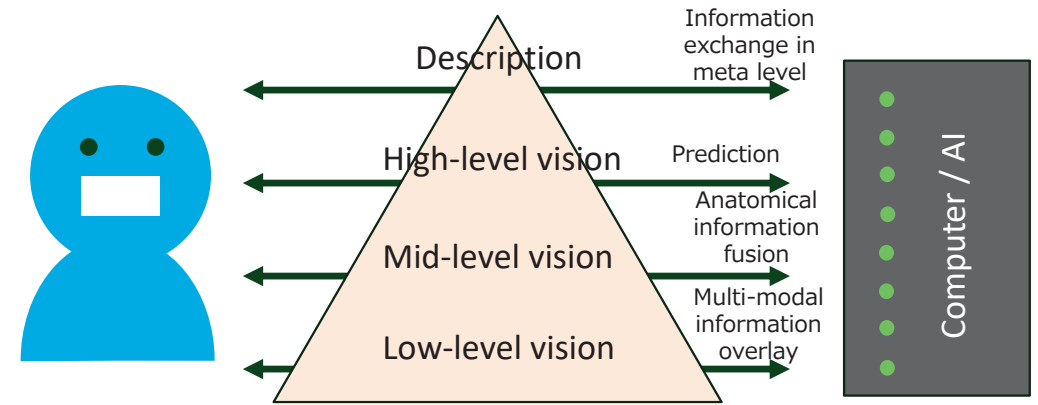
# 1974.11のコード



## Abstraction level in medical assistance



## Abstraction level in medical assistance



## コンピュータ支援画像診断

### • 当初は自動診断

– Early studies on quantitative analysis of medical images by computer<sup>18, 19, 20, 21, 22, 23</sup> were reported in the 1960s. At that time, it was generally assumed that computers could replace radiologists in detecting abnormalities, because computers and machines are better at performing certain tasks than are human beings. Thus, **the concept of computer diagnosis or automated diagnosis in radiology was established at that time.**

### • 途中で少し路線変更し、コンピュータは医師のセカンドオピニオンとして機能する「コンピュータ支援診断」へと

– *In the 1980s, however, another approach emerged which assumed that **the computer output could be utilized by radiologists, but did not replace them.** [K. Doi 2004]*

Kunio Doi, Overview on research and development of computer-aided diagnostic schemes, Seminars in Ultrasound, CT and MRI, Volume 25, Issue 5, 2004, 404-410, <https://doi.org/10.1053/j.sult.2004.02.006>.



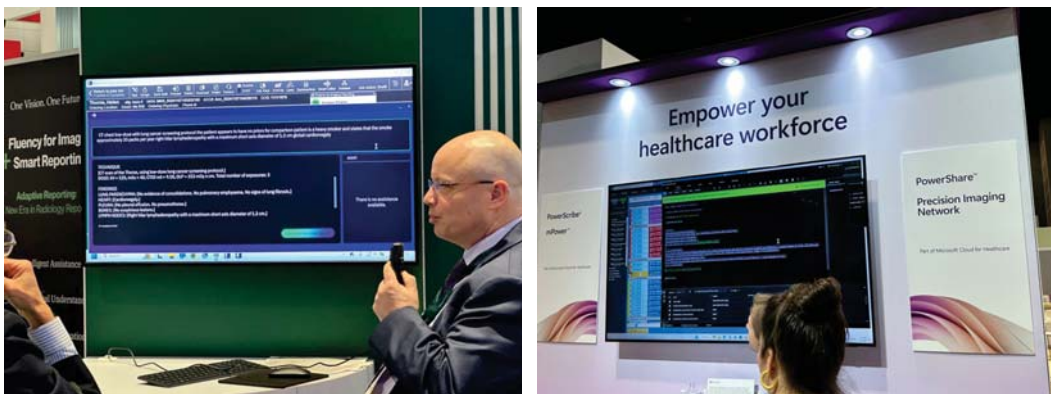
## Large Language Model (LLM)による総合的な診断

- 大規模言語モデル (LLM) とは
  - 人間の言語を理解し、生成するように設計されたAIモデル
  - 主なアーキテクチャは Transformer
- 「大規模 (Large)」の意味
  - モデル内部に 数十億規模のパラメータ を持つ
  - 大量のデータと計算資源を必要とする
- 事前学習 (Pre-training)
  - 書籍・記事・ウェブサイトなどの大規模テキストで学習
  - 学習期間は 数週間~数か月
- 学習の仕組み
  - 「次に来る単語」を予測しながら言語を学習
  - 文法、事実、一般知識、文体などを獲得
- 学習の段階
  - 自己教師あり学習 (Self-supervised learning) → 言語の基本構造を理解
  - 命令調整 (Instruction tuning) → 質問応答や指示への応答能力を獲得
- 画像特徴をLLMに入力するとLLMが画像に関する**網羅的な解釈**を文章として返す

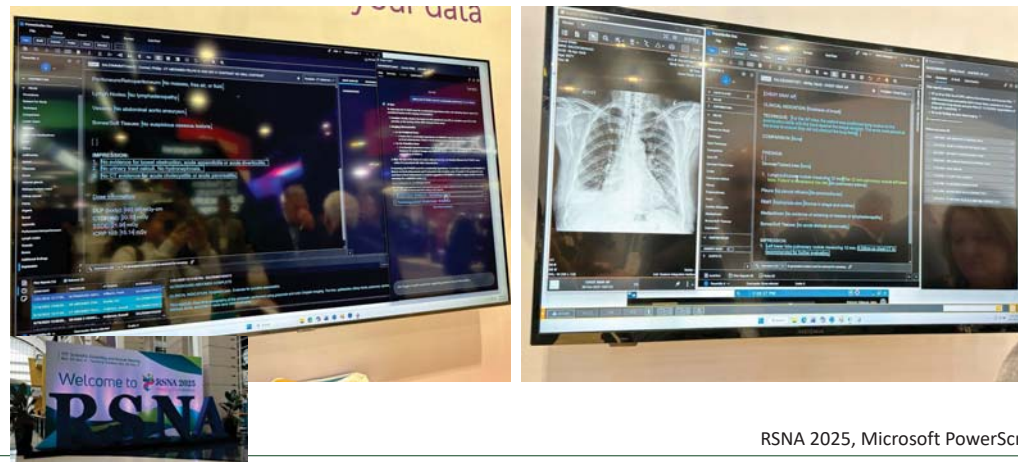
### 代表的なLLMの例:

- GPTシリーズ (OpenAI)
- Claude (Anthropic)
- Gemini (Google DeepMind)
- LLaMA (Meta)
- Mistral (Mistral AI)
- Qwen (Mistral AI)

## 言語モデルを利用した読影レポート作成支援 RSNA2024

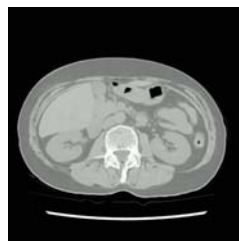


## DictationされたFindingsからImpressionを自動生成



## 放射線分野での自動診断

- 放射線診断レポートの作成
  - 放射線診断の過程で放射線レポートが作成
  - FindingsとImpressions
- 大規模言語モデル (LLM) の発展
  - 自然言語処理のためのLLM
  - AIエージェントシステムの基盤
- Vision – Language Model (VLM)
  - 視覚と言語を統合的に扱うモデル
  - VLM は、画像とそれに関連する質問を入力として受け取り、質問に対する回答を出力



(Findings) Bilateral pleural effusion is unchanged from last time. There is mild collapse of the left lung base. A nodule in the central left lower lobe is unchanged but gradually increasing. A small localized ground-glass opacity in the left lower lobe is also unchanged. Right pleural thickening shows no significant change. Small scattered nodules in both lungs show no growth and are likely benign post-inflammatory. There is no significant lymph node enlargement. The right mediastinal soft tissue shadow is unchanged. Bilateral renal cysts are present, with no hydronephrosis or ascites

## ARRG – Automated Radiology Report Generation

368 IEEE REVIEWS IN BIOMEDICAL ENGINEERING, VOL. 18, 2025

**Automated Radiology Report Generation: A Review of Recent Advances**

Phillip Sloan, Philip Clatworthy, Edwin Simpson, and Majid Mirmehdi

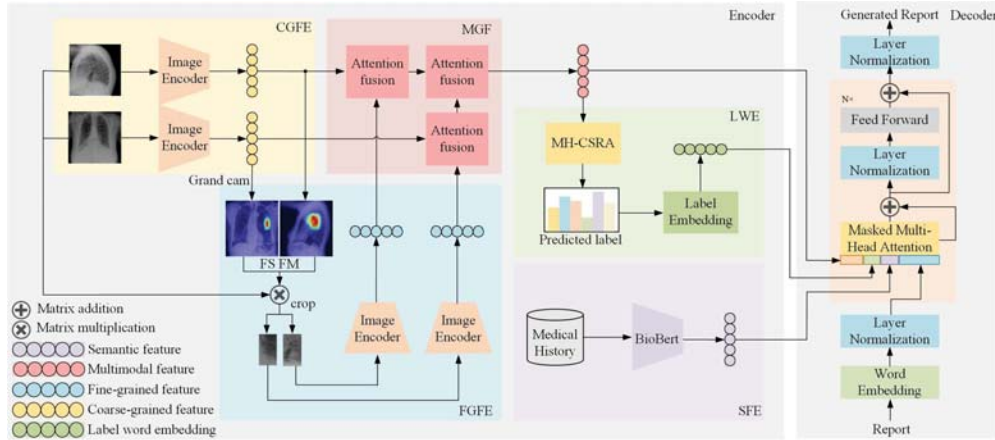
**Abstract**—Increasing demands on medical imaging departments are taking a toll on the radiologist’s ability to deliver timely and accurate reports. Recent technological advances in artificial intelligence have demonstrated great potential for automatic radiology report generation (ARRG), sparking an explosion of research. This survey paper conducts a methodological review of contemporary ARRG approaches by way of (i) assessing datasets based on characteristics, such as availability, size, and adoption rate, (ii) examining deep learning training methods, such as contrastive learning and reinforcement learning, (iii) exploring state-of-the-art model architectures, including variations of CNN and transformer models, (iv) outlining techniques integrating clinical knowledge through multimodal inputs

**TABLE I**  
RECENT ARRG REVIEW ARTICLES AND THEIR COVERAGE INCLUDING TOTAL NUMBER OF CITATIONS AND THE TOTAL NUMBER OF ARTICLES RELATED TO ARRG OR MEDICAL REPORT GENERATION

Author	Year	Coverage	# Cited	# ARRG
Allaoui et al. [16]	2018	2015-2018	50	10
Pavlopoulos et al. [17]	2019	2015-2019	65	7
Monshi et al. [12]	2020	2015-2019	105	14
Kaur et al. [14]	2022	2015-2021	83	19
Mesina et al. [15]	2022	2016-2021	167	40
Liao et al. [13]	2023	2016-2021	154	41
Our Review	2023	2020-2023	154	83



## X線画像からの診断レポート自動生成



Junze Fang, Suxia Xing, Kexian Li, Zheng Guo, Ge Li, Chongchong Yu, Automated generation of chest X-ray imaging diagnostic reports by multimodal and multi granularity features fusion, Biomedical Signal Processing and Control, 2025, <https://doi.org/10.1016/j.bspc.2025.107562>



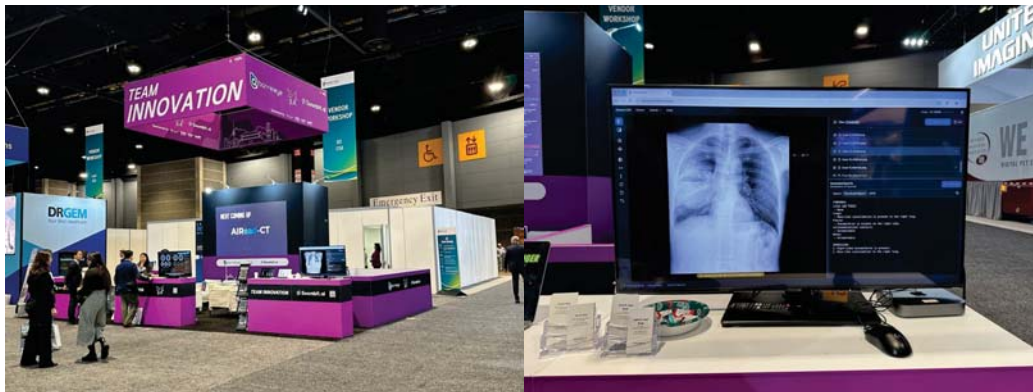
## X線画像からの診断レポート自動生成

Image	Ground Truth	Base	MMG
	the cardiomeastinal silhouette is within normal limits the lungs are normally inflated without evidence of focal airspace disease, pleural effusion or pneumothorax, no acute osseous abnormality. Large hiatal hernia.	the cardiomeastinal contours are within normal limits, the lungs are clear bilaterally specifically, no evidence of focal consolidation pneumothorax or pleural effusion, no acute bone abnormality.	the cardiomeastinal silhouette is within normal limits for size and contour, the lungs are normally inflated without evidence of focal airspace disease, pleural effusion or pneumothorax, no acute bone abnormality.
	there is moderate cardiomegaly, there are bilateral interstitial opacities increased since the previous exam, no focal airspace consolidation pleural effusions or pneumothorax, no acute bony abnormalities.	there is moderate cardiomegaly, no pleural effusion, pneumothorax or focal airspace disease, no acute bony abnormality.	there is moderate cardiomegaly, there is no focal airspace consolidation pleural effusions, no visualized pneumothorax, no acute bony abnormality.
	normal heart size, stable tortuous thoracic aorta prior granulomatous disease healed rib fractures appear, stable focal opacity is noted in the left midlung overlying the 9th posterior rib which represents healing rib callus, no pneumothorax or pleural effusion.	the heart is normal in size, the mediastinum is stable, there is no pleural effusion or pneumothorax, aorta are tortuous, focal opacity is noted, arthritic changes of the skeletal structures are noted.	heart size within normal limits, stable mediastinal contours, the mediastinum is stable, no pleural effusion or pneumothorax, there are stable tortuous of the aorta, focal opacity is noted in the left.
	heart size and pulmonary vascularity appear within normal limits, there is mild tortuosity to the descending thoracic aorta, the lungs are free of focal airspace disease, no pleural effusion or pneumothorax is seen, no discrete nodules or adenopathy, degenerative changes are present in the spine.	the heart size is normal, the lungs are free of focal airspace disease, no pleural effusion or pneumothorax, degenerative changes are in the spine.	heart size and pulmonary vascularity appear within normal limits, the lungs are free of focal airspace disease, no pleural effusion or pneumothorax is seen, no discrete nodules, degenerative changes of the thoracic spine.

Junze Fang, Suxia Xing, Kexian Li, Zheng Guo, Ge Li, Chongchong Yu, Automated generation of chest X-ray imaging diagnostic reports by multimodal and multi granularity features fusion, Biomedical Signal Processing and Control, 2025, <https://doi.org/10.1016/j.bspc.2025.107562>



## 胸部X線写真の自動診断

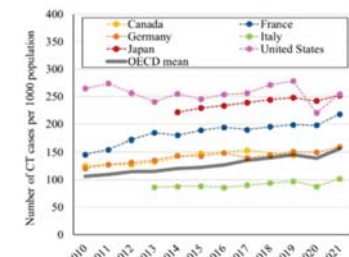


Soombit.ai



## CT検査数の増大

- 毎年何百万件ものCT検査が実施
  - CT検査の自動解析
- 経時解析は臨床上極めて重要
  - 以前のCTと比較治療効果の評価や経過観察
  - 複数CTの読影は時間と労力がかかる
  - 認知負荷・疲労・ストレス増大



人口千人当たりのCT検査数[1]

自動レポート生成システムの必要性が高い

[1] Aoyama, T., et al.: A cross-national investigation of ct, mri, pet, mammography, and radiation therapy resources and utilization (2024)

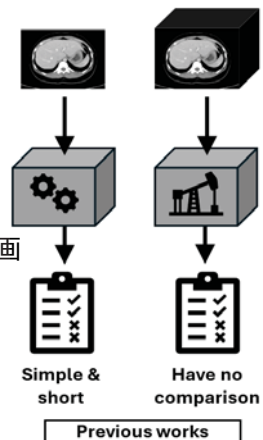
## CT画像と所見文



- [腹部CT]肝：pneumobiliaあり。脾：n.p.胆嚢：摘出後。膵：膵頭十二指腸切除術後。残膵はやや萎縮。腎：両腎に嚢胞を認めます。副腎：n.p.リンパ節腫大なし。腹水なし。

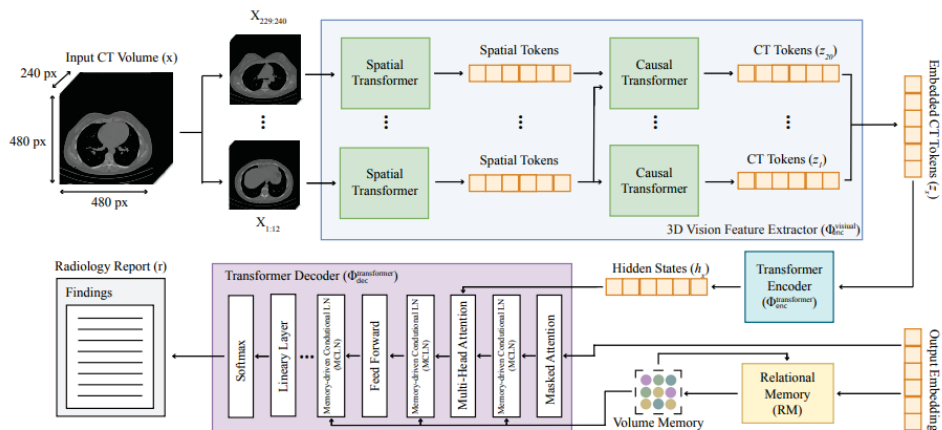
## 医用画像LMM

- 2次元画像中心
  - VLMの多くが2D画像に特化
  - ボリュームとして情報を捉えられない
  - 2次元モデルの3次元モデルへの拡張は不十分
- 3次元画像を対象としたものは少ない
  - 過去画像の情報をを用いるものは極めて少ない
  - 一部研究 (CT2RepLong[3]) は関連性記憶を用いて経時画像に対応
  - 日本語所見文等を用いて学習させたものはない
  - 日本での所見文と他国での所見文にも差
  - **学習用データも大量**に必要



[2] Blankemeier, L., et al.: Merlin: A vision language foundation model for 3d computed tomography (2024)  
 [3] Hamamci, I.E., et al.: CT2Rep: Automated radiology report generation for 3d medical imaging (2024)

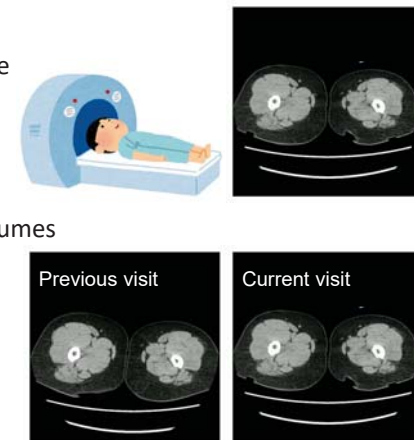
## Various-vision language models CT2Rep<sup>4)</sup>



4. Ibrahim Ethem Hamamci, et al., CT2Rep: Automated Radiology Report Generation for 3D Medical Imaging, <https://doi.org/10.48550/arXiv.2403.06801>

## CT image interpretation

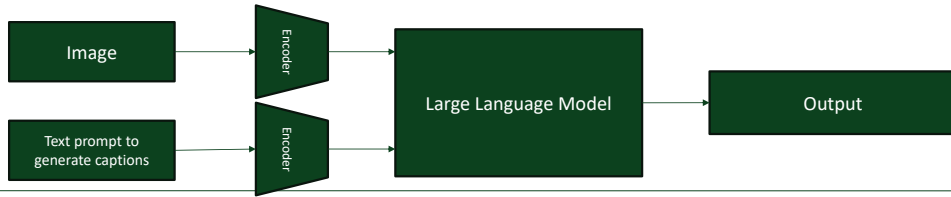
- Single-shot image interpretation
  - Diagnosis has been made only from one volume
- Longitudinal image interpretation
  - Diagnosis has been made from longitudinal volumes to find changes





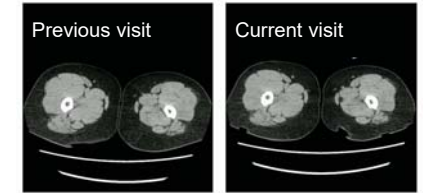
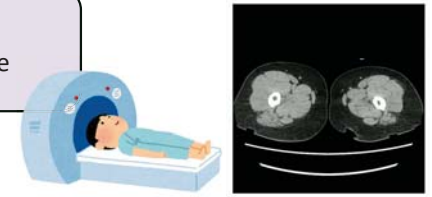
## Vision-Language Model (VLM)

- 画像と言語 (テキスト) の両方を扱うAIモデル
- 画像から キャプション (説明文) を生成可能
  - 説明文の代わりにロボット操作コマンドも生成可能 (Vision Language Action)
- 放射線診断レポート作成補助、CT/MR画像の自動生成支援に応用可能
- 一般的にVLMは以下の要素で構成
  1. 画像エンコーダ (Image Encoder)
  2. テキストエンコーダ (Text Encoder)
  3. 言語モデル (Vision-Language Model)
  4. デコーダ (Decoder)



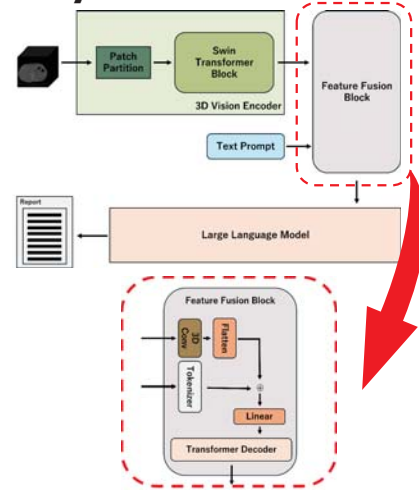
## CT image interpretation

- Single-shot image interpretation
  - Diagnosis has been made only from one volume
- Longitudinal image interpretation
  - Diagnosis has been made from longitudinal volumes to find changes

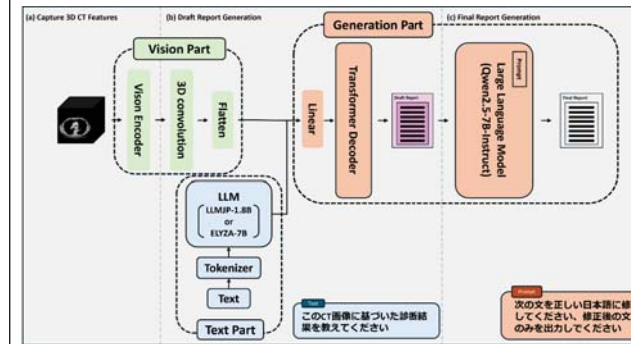


## CT report generation by VLM

- **Using a 3D encoder to directly process 3D CT volumes**
  - Visual Encoder
    - pretrained 3D Swin Transformer is used to extract features from 3D CT volumes
  - **Feature Fusion**
    - A 3D convolutional layer followed by flattening merges spatial information
  - Language Model
    - A Japanese LLM (ELYZA-japanese-Llama-2-7b) processes prompts and generates the final report
  - End-to-End Flow
    - 3D CT volume → Feature extraction → Transformer Decoder → Prompted LLM → Japanese report



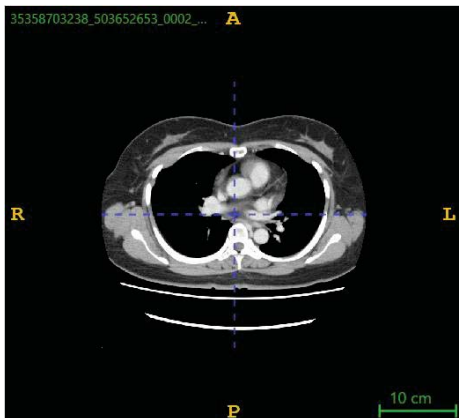
## CT report generation by VLM



- Capture 3D CT Features
  - Utilize a pretrained 3D Swin Transformer to capture CT features from 3D CT volumes.
- Draft Report Generation (Stage 1)
  - CT features are first transformed into the text embedding space using a 3D convolutional layer followed by a flattening layer.
  - The text prompt is transformed into text embeddings through a pre trained LLM (ELYZA-7B and LLMJP-1.8B).
  - Transformed CT features are concatenated with the text embeddings and the concatenated features are sent to the Transformer Decoder Layer to generate a draft report.
- Final Report Generation (Stage 2)
  - The LLM (Qwen2.5-7B-Instruct) refines the draft report guided by prompts to produce the final report.



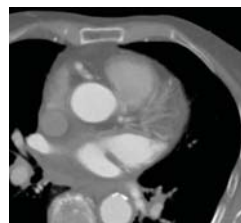
## Result



Output	Ground Truth
腎に嚢胞が認められるが、肝、脾、胆嚢、膵、副腎には異常なし。リンパ節腫大も認められず、腹水はなし。 (Cysts are observed in the kidney, but no abnormalities are found in the liver, spleen, gallbladder, pancreas, or adrenal glands. There is no lymphadenopathy, and no ascites is present.)	両腎に嚢胞を認めるが、肝、脾、胆嚢、膵、副腎には異常なし。リンパ節腫大も見られず、腹水はなし。 (Cysts are observed in both kidneys, but no abnormalities are found in the liver, spleen, gallbladder, pancreas, or adrenal glands. There is no lymphadenopathy, and no ascites is present.)

## E-2 冠動脈造影3D CT像 (目的特化3D CT)に対するレポート生成

症例1 (入力は3次元CT)



(AI生成) 気管は開いています。縦隔内に病的なリンパ節腫脹は認められませんでした。心臓の検査では、左前降支 (LAD) と右冠動脈が石灰化した粥状硬化斑点で塞栓性狭窄を示しています。心拡大あり。両側胸腔において胸水や肥厚は認められませんでした。上腹部の評価では、肝実質疾患に関連する可能性のある低密度領域が観察されました。骨構造には変形があります。椎体後部要素および脊髄圧迫症候群が見られます。

症例2 (入力は3次元CT)

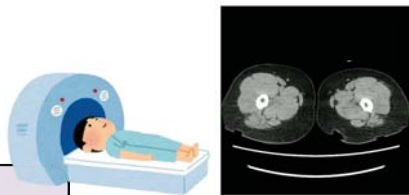


(AI生成) 気管、主支気管は開いています。右肺上葉に軽度の線状無気肺が見られます。両側肺下部には胸膜-実質後遺症があります。左肺下葉前内基底セグメントでは、直径約10mmの結節性病変 (非特異的) と疑わしい腫瘤様所見があり、その周囲でわずかなすりガラス密度が観察されます。この外観から、感染や炎症過程による可能性があると考えられました。臨床および検査との相関が必要です。心拡大あり。大動脈及び冠動脈には石灰化したアテロームプラークが認められます。縦隔リンパ節群において病理学的に拡大したリンパ節は確認されませんでした。食道壁厚の増加も認められず、胃内容物は自然なものです。胆嚢腔内には石像のような高密度領域が存在します。骨構造には破壊的な溶骨性または硬化性占位病変は認められませんでした。

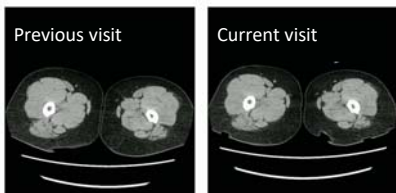


## CT image interpretation

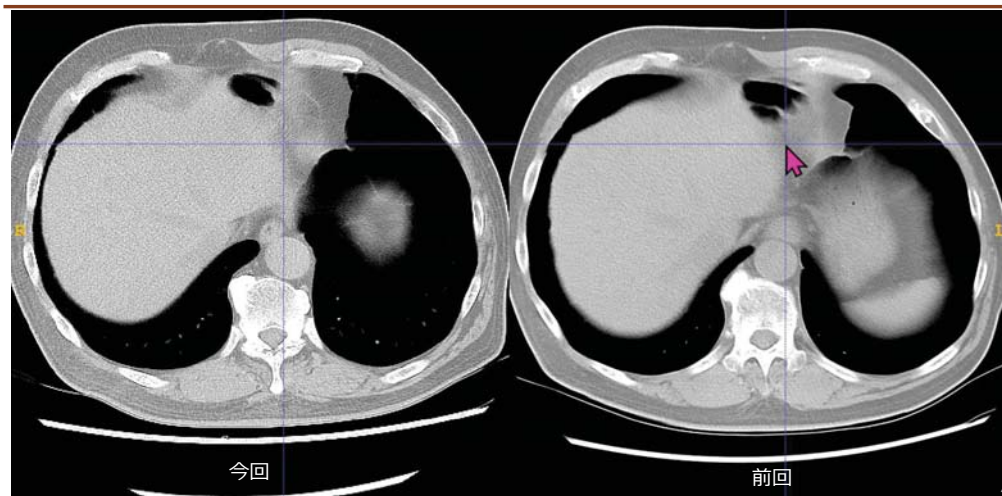
- Single-shot image interpretation
  - Diagnosis has been made only from one volume



- Longitudinal image interpretation
  - Diagnosis has been made from longitudinal volumes to find changes

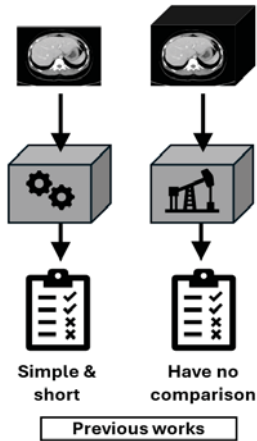


## 比較読影 - 2つの画像を比較しながら画像を読む



# 医用画像LMM

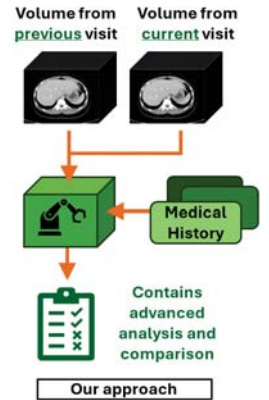
- 2次元画像中心
  - VLMの多くが2D画像に特化
  - ボリュームとして情報を捉えられない
  - 2次元モデルの3次元モデルへの拡張は不十分
- 3次元画像を対象としたものは少ない
  - 過去画像の情報を用いるものは極めて少ない
  - 一部研究 (CT2RepLong[3]) は関連性記憶を用いて経時画像に対応
  - 日本語所見文等を用いて学習させたものはない
  - 日本での所見文と他国での所見文にも差
  - **学習用データも大量**に必要



[2] Blankemeier, L., et al.: Merlin: A vision language foundation model for 3d computed tomography (2024)  
 [3] Hamamci, I.E., et al.: CT2Rep: Automated radiology report generation for 3d medical imaging (2024)

# 経時CT画像の読影レポート生成

- 経時CT画像を対象とした読影レポート生成
  - 2つの3次元CT画像から読影レポートを生成
- 視覚言語モデル (VLM)により実現
  - 経時3次元CT画像間の変化を記述
  - 以前のデータと介入情報(例: 胃半摘出術後など)の利用
  - 日本語でのレポート生成
    - 日本の臨床現場で記述されているレポートと同様なレポート

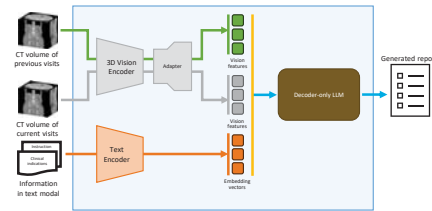


# LMMによる経時画像読影レポート自動生成

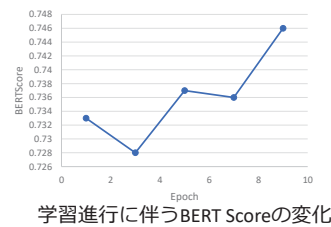
統合ヘルスケアシステム

名古屋大学

- 経過観察のために撮影された2つの3次元X線CT画像を解析し、自然な日本語所見文を生成する世界初の生成AI
- 経過観察のために2つの3次元CT画像と所見文を生成するための質問文を入力すると所見文を生成
- 大規模医用画像データベースの3次元CT画像と付随所見文を用いて訓練
  - 3.5万ペア (7万件) の3次元CT画像 (約700万断面画像)
  - 1万4千人の経時画像
- 追加質問にも対応可能
- 2025/4/24にプレスリリース



モデルの概要

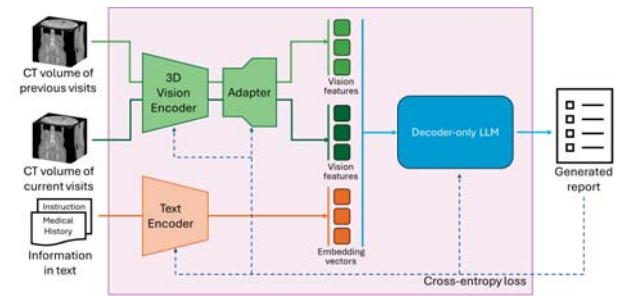


学習進行に伴うBERT Scoreの変化



# モデル構造

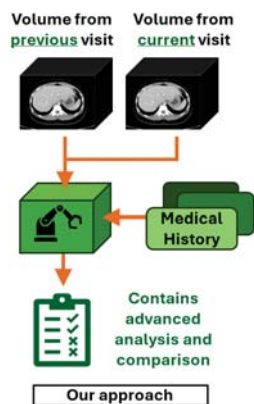
- モデル入力
  - 過去と現在のCT画像
  - 比較読影レポート
- モデル出力
  - 所見文
- カギとなるコンポーネント
  - 3D Vision Encoder
  - Adapter
  - Text Encoder
  - Vision-Language Decoder



# データセットとモデル

## 国立情報学研究医療ビッグデータ研究センターDBを利用

- 14,000 症例 (M:F=8:6)
- 50,000 3D-CT 画像
  - Size on axial plane 512×512 pixel
- 対応所見文
- 35,000経時画像ペア
  - 2つの3次元画像と所見文を入力
- モデル
  - Vision-encoder SwinTransformer
  - Decoder-only LLM: Gemma-2-9B-it



# 経時3次元CT像の自動レポートニング

## Press Release



配布先: 文部科学記者会、科学記者会、名古屋教育記者会

2025年4月24日

報道機関 各位

世界初、CT 画像を AI が比較・解析し日本語の所見文を生成  
～放射線科医の経過観察支援に期待～

### 【研究概要】

名古屋大学大学院情報科学研究科の森 健策 研究室、情報・システム研究機構 国立情報学研究所医療ビッグデータ研究センターの共同研究グループは、経過観察のために撮影された2つの3次元X線CT画像<sup>1)</sup>を解析し、自然な日本語所見文を生成する世界初の生成AI<sup>2)</sup>を開発しました。本生成AIを利用し、2つの異なるCT画像を比較して、その結果を日本語文章として出力することができるようになりました。これによって、放射線科医の正確なCT画像読影所見文の作成を支援することが期待されます。

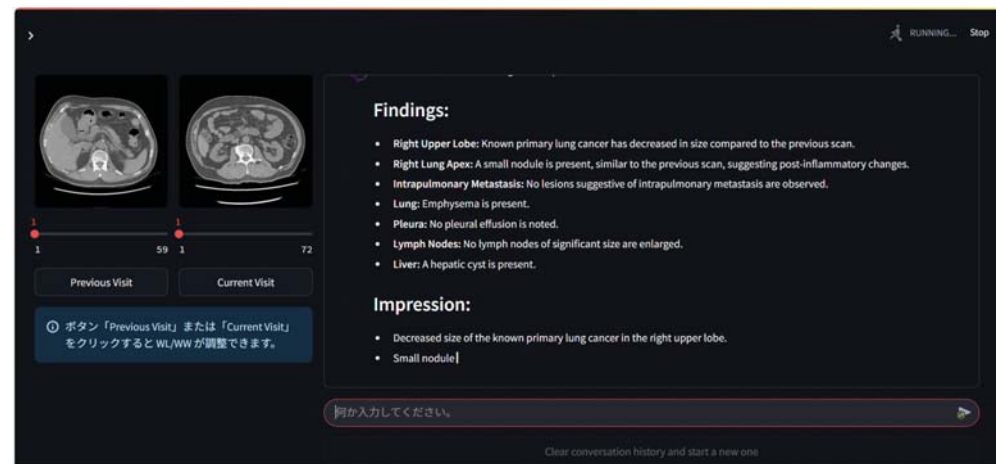
本研究成果は、2025年4月30日に開催のSIP第3期「統合型ヘルスケアシステムの構築における生成AIの活用」公開シンポジウムで発表されます。



# 経時画像読影レポート自動生成

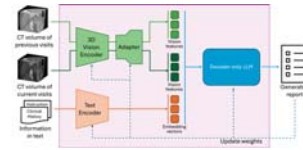


# 単なる所見文生成だけでなく指示に応じたレポート整形



# 経時3D CT画像所見文生成モデルベンチマーク

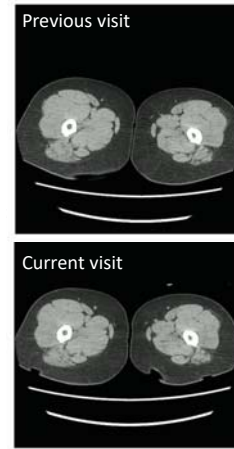
- SIGMA+
  - SIP E-2 名大Gを中心に開発された経時3D CT画像所見文生成モデル
- CT2RepLong<sup>[1]</sup>と比較
  - CT2RepLongはNII DBを用いて名大にて再学習
  - 経時3D CT画像所見文生成モデルとして唯一比較可能なモデル
- 全てのスコアにおいてCT2RepLong<sup>[1]</sup>を上回る



	BLEU-1	BLEU-2	BLEU-3	BLEU-4	METEOR	ROUGE-L
CT2RepLong	0.298	0.216	0.174	0.144	0.276	0.235
SIGMA	<b>0.331</b>	<b>0.248</b>	<b>0.199</b>	<b>0.162</b>	<b>0.312</b>	<b>0.253</b>

[1] Hamami, I.E., et al.: CT2Rep: Automated radiology report generation for 3D medical imaging (2024)

# レポート生成例



入力3次元画像

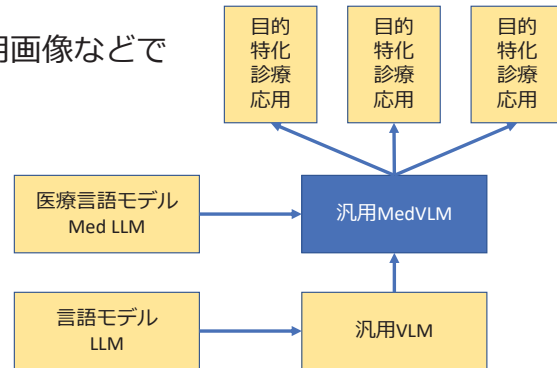
	Text	Text (Translated to English)
プロンプト	所見文を作成してください。直腸癌術後。	Please create a report. Postoperative rectal cancer.
AI出力	明らかな局所再発は指摘できません。有意なサイズのリンパ節腫大は指摘できません。明らかな肝転移は指摘できません。肝嚢胞あり。水腎症なし。腹水なし。明らかな肺転移は指摘できません。胸水なし。	No obvious local recurrence can be identified. No significant lymph node enlargement can be identified. No clear liver metastasis can be identified. Liver cysts are present. No hydronephrosis. No ascites. No clear lung metastasis can be identified. No pleural effusion.
放射線科医によるレポート	明らかな局所再発なし。両肺に多発小結節影あり。サイズの増大ありません。一部の結節はリング状に内部欠損みられます。新規結節は指摘されません。有意なサイズのリンパ節腫大はみられません。胸水なし。甲状腺両葉のLDAに大きな変化なし。明らかな肝転移なし。肝嚢胞あり。CVポートデバイスあり。その他、特記所見なし。	No obvious local recurrence. Multiple small nodules in both lungs, with no increase in size. Some nodules show ring-like internal defects. No new nodules identified. No significant lymph node enlargement. No pleural effusion. No significant changes in the low-density areas (LDA) of both thyroid lobes. No clear liver metastasis. Liver cysts are present. CV port device is present. No other notable findings.

Similar to the findings written by radiologist

K. Nguyen, C. Wang, M. Oda, K. Mori. "Automated findings report generation using VLM from longitudinal 3D CT volumes," CARS 2025 Computer Assisted Radiology and Surgery 39th International Congress and Exhibition

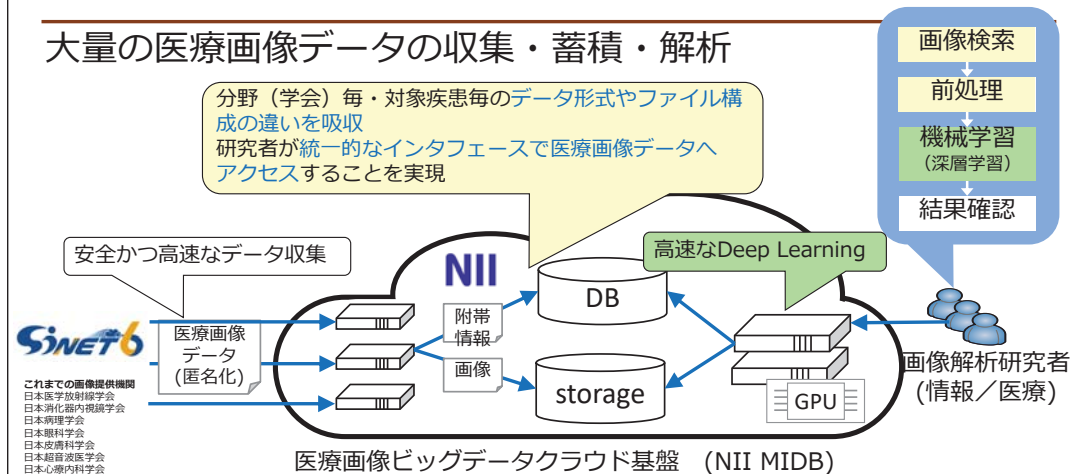
# 医用画像VLM位置づけ

- 医用画像を解釈して文章などを生成できる
- 汎用的なモデルをベースに医用画像などでさらに訓練を行う
- 言語モデルも重要な役割
  - 言語モデルは誰かが構築したものを再学習させることが多い
  - 海外製モデルが現在主流



# 医療画像ビッグデータクラウド基盤 - 蓄積されるのは日々のデータ

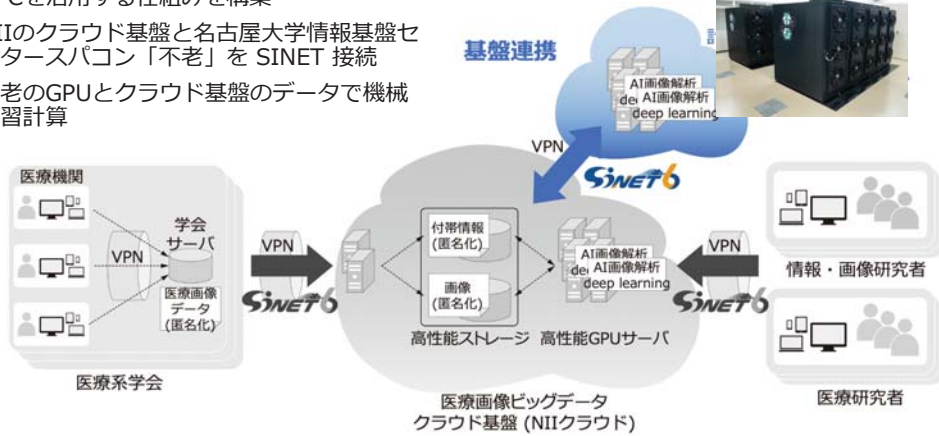
## 大量の医療画像データの収集・蓄積・解析



医療画像ビッグデータクラウド基盤 (NII MIDB)  
2017年にSINETを活用した画像収集クラウド基盤を構築

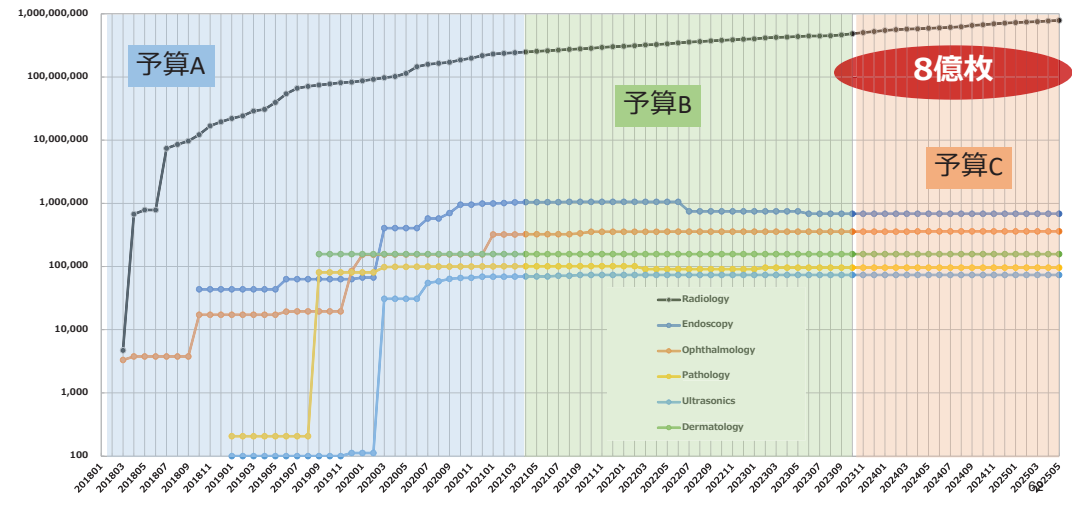
## DB-GPU基盤連携

- 計算資源の逼迫に対処するために、外部のHPCを活用する仕組みを構築
- NIIのクラウド基盤と名古屋大学情報基盤センタースパコン「不老」をSINET接続
- 不老のGPUとクラウド基盤のデータで機械学習計算

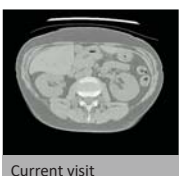
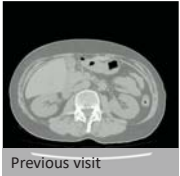


61

## クラウド基盤に格納された画像数 2025年5月末現在



## 生成AIによる診断文をどのように評価するか？



**Expert:** 両側胸水は前回同様。左肺底部の肺虚脱程度あり。左下葉中極側の結節は前回同様、経時的に増大あり。左下葉に小さな限局性すりガラス影は前回同様です。右胸膜肥厚に著変ありません。その他、両肺に散在する小結節影に増大傾向なく、炎症後変化など良性結節疑い。有意なサイズのリンパ節腫大なし。右縦隔の軟部影は前回同様です。両腎嚢胞あり。水腎症なし。腹水なし。  
 (Bilateral pleural effusion is unchanged from last time. There is mild collapse of the left lung base. A nodule in the central left lower lobe is unchanged but gradually increasing. A small localized ground-glass opacity in the left lower lobe is also unchanged. Right pleural thickening shows no significant change. Small scattered nodules in both lungs show no growth and are likely benign post-inflammatory. There is no significant lymph node enlargement. The right mediastinal soft tissue shadow is unchanged. Bilateral renal cysts are present, with no hydronephrosis or ascites.)

**Input:** 所見文を作成してください。  
 (Please generate a report.)

**SIGMA:** 右肺上葉の結節影は前回同様です。両肺に慢性気道炎症疑いあり。前回同様です。胸水なし。有意サイズのリンパ節腫大なし。  
 (Nodules in the right upper lobe are unchanged. Both lungs show findings suggestive of chronic airway inflammation, unchanged from last time. No pleural effusion. No significant lymph node enlargement.)

**Input:** 所見文を作成してください。右肺上葉癌術後、特発性乳び胸術後  
 (Please generate a report. Post-right upper lobe lung cancer surgery, post-idiopathic chylothorax surgery)

**SIGMA+:** 右肺尖部には浸潤影、すりガラス影あり。前回より増大。肺炎の疑い。右肺上葉の結節影は前回同様です。右肺尖部の濃厚影は前回同様です。背景肺に気腫性変化あり。左胸水は前回同様です。有意なサイズのリンパ節腫大なし。  
 (In the right lung apex, there is an infiltrative and ground-glass shadow, increased since last time, suggesting possible pneumonia. Nodules in the right upper lobe are unchanged. Dense shadow in the right lung apex is unchanged. Emphysematous changes are seen in the background lung. Left pleural effusion is unchanged. No significant lymph node enlargement.)

Note: Original input and output are in Japanese, the English version was translated by authors, written in the parentheses.  
 Green for findings match with the expert  
 Red for findings differ from the expert

(名大 情報学 森健策)

## 医療機器評価における生成AI利用

- 自然言語処理の場合の代表的なメトリック
  - BLEU-n: Based on n-gram matches
  - METEOR: Considers synonyms, stemming, and word order
  - ROUGE-L: Measures longest common subsequence
  - BertScore: Measures semantic similarity
  - $r_m$  (Matching Rate): Fraction of output findings that match the expert's findings
  - GREEN: Evaluate factual correctness and clinical relevance
- 上記の評価メトリックは生成された診断文の正しさを表すか？
  - GPTを用いた評価も行われる
  - プロンプトエンジニアリングで評価



## 評価結果の例

Method	Natural Language Generation							Clinical Efficacy	
	BLEU-1	BLEU-2	BLEU-3	BLEU-4	METEOR	ROUGE-L	BERTScore	$r_m$	GREEN
W/o clinical history (SIGMA)	0.313	0.243	0.183	0.147	0.230	0.253	0.730	0.306	0.272
W/ clinical history (SIGMA+)	0.318	0.247	0.196	0.153	<b>0.322</b>	<b>0.287</b>	<b>0.746</b>	<b>0.353</b>	<b>0.361</b>
Below shows performance shown in Ref. [3] as reference. c.f. CT2RepLong is English model, and our model is Japanese model, direct comparison is impossible.									
CT2RepLong[3]	<b>0.374</b>	<b>0.327</b>	<b>0.304</b>	<b>0.401</b>	0.285	0.263	-	-	-

- BLEU-n: Based on n-gram matches
- METEOR: Considers synonyms, stemming, and word order
- ROUGE-L: Measures longest common subsequence
- BertScore: Measures semantic similarity
- $r_m$  (Matching Rate): Fraction of output findings that match the expert's findings
- GREEN: Evaluate factual correctness and clinical relevance

[3] Hamamci, I.E., et al.: CT2Rep: Automated radiology report generation for 3d medical imaging (2024)



## BLEUスコアの仕組み (機械翻訳の評価指標)

**1. 参考訳 (正解) と候補訳を比較**

参考訳 (Reference): "The cat is sitting on the mat."  
 候補訳 (Translation): "The cat sits on a mat."

**2. n-gramの一致度を計算**

1-gram: The, cat, is, sitting, on, the, mat.  
 2-gram: The cat, cat sits, sits on, on a, a mat.  
 3-gram: The cat sits, cat sits on, sits on a, on a mat.  
 4-gram: The cat sits on, cat sits on a, sits on a mat, cat sits on a mat.

**3. n-gram通合率 (Precision) を計算**

1-gram通合率 =  $\frac{\text{一致数}}{\text{総n-gram数}} = \frac{4}{6}$   
 2-gram通合率 =  $\frac{3}{5}$   
 3-gram通合率 =  $\frac{2}{3}$   
 4-gram通合率 =  $\frac{1}{2}$

BLEU = BP × exp( $\sum w_n \log P_n$ )  
 ( $w_n$ : 重み,  $P_n$ : n-gram通合率)

**4. 短さペナルティ (BP)**

候補訳が短すぎる場合に調整  
 参考訳: "The cat is sitting on the mat."  
 候補訳: "The cat sits on a mat."  
 短い → ペナルティ  
 → BP < 1 でスコア調整

今回のBLEU ≈ 0.75  
 (1に近いほど良い翻訳)



## 文章の評価例 BLEUスコア

- 文例1
  - A) 明日雨が降るので傘を持っていくとよいです。
  - B) 明日は雨なので傘を持っていきましょう。
  - BLEU-4 0.3763

← ほぼ同じ意味でもスコア低い
- 文例2
  - A) 明日雨が降るので傘を持っていくとよいです。
  - B) 今日日は雨なので傘を持っていきましょう。
  - BLEU-4 0.3561

← 今日か明日が違うしスコアも低い
- 文例3
  - A) 右肺上葉に結節影が認められます
  - B) 左肺上葉に結節影が認められます
  - BLEU-4 0.9254

← 左と右を間違える致命的誤りにも関わらずスコアが高い



## 評価における生成AIの利用

単語レベルだけの評価では難しい部分も評価

### プロンプト

以下はコンピュータによって生成された診断文と専門医が書いた診断文です。2つの診断文の一致度を評価して0-100のスコアで返してください。

**専門医** 胃体中部後壁の既知の胃癌は縮小を維持しているようです。以前に腫大を認めた胃小弯リンパ節、傍大動脈リンパ節はいずれも縮小を維持しています。両肺に5mm大までのすりガラス結節や粒状影を複数認めます。前回とほぼ著変ありません。多発肝嚢胞あり。右腎嚢胞あり。胆道系、脾臓、副腎、睪臓に明らかな異常なし。子宮筋腫と思われる腫瘤を複数認めます。骨盤底に腹水貯留あり。骨転移を疑う変化は指摘できません。  
**コンピュータ** 胃壁の肥厚は改善しています。胃幽門部周囲のリンパ節や播種は縮小維持しています。他、明らかに増大傾向なリンパ節は指摘できません。多発肝転移は前回より全体に縮小しています。腹水は減少しています。両肺に散見される小結節影は前回同様です。胸水なし。

### 一致度スコア: 45 / 100

胃病変の評価専門医: 胃体中部後壁の既知の胃癌は「縮小を維持」コンピュータ: 胃壁の肥厚は「改善」 → どちらも「病変の縮小・改善」という方向性で一致。  
 リンパ節の評価専門医: 胃小弯リンパ節・傍大動脈リンパ節「縮小を維持」コンピュータ: 胃幽門部周囲リンパ節や播種は「縮小維持」、増大リンパ節なし  
 → 部位の表現は少し違いますが、「リンパ節は縮小維持で、新たな増大なし」でほぼ一致。

**NAGOYA UNIVERSITY**

## FDA Request for Public Comment: Measuring and Evaluating AI-enabled Medical Device Performance in the Real-World

**Measuring and Evaluating Artificial Intelligence-enabled Medical Device Performance in the Real-World**  
Request for Public Comment

This Request for Public Comment is intended for discussion purposes only and does not represent draft or final guidance. It is not intended to propose or implement policy changes regarding the evaluation of devices which integrate artificial intelligence (AI), including generative AI (GenAI)-enabled technology. This document is not intended to communicate the FDA's proposed (or final) regulatory expectations but is instead meant to seek early feedback from groups and individuals outside the Agency and advance a broader discussion among the AI healthcare ecosystem on this topic.

The objective of this 'Request for Public Comment' is to obtain comment and feedback from interested parties and the public on a series of questions related to the current, practical approaches to measuring and evaluating the performance of AI-enabled medical devices in the real world, including strategies for identifying and managing performance drift, such as detecting changes in input and output. Please submit your comments to <https://www.regulations.gov>, Docket No. FDA-2025-N-4203 for 'Measuring and Evaluating Artificial Intelligence-enabled Medical Device Performance in the Real World, Request for Public Comment.' FDA intends to consider all comments timely submitted to this docket (FDA-2025-N-4203) by December 1, 2025, related to this topic.

**Background**  
AI, including GenAI, presents opportunities to improve patient outcomes, advance public health, and accelerate medical innovation. At the same time, these technologies introduce new considerations as it relates to assuring the maintained safety and effectiveness of AI-enabled medical devices across the total product life cycle, and particularly with respect to assessing their performance, safety, and reliability after deployment in real-world settings.

The U.S. Food and Drug Administration (FDA or the Agency) is seeking information from interested parties and the public on best practices, methodologies, and approaches for measuring and evaluating real-world performance of AI-enabled medical devices. This includes

**パフォーマンス評価の重要性**  
AI医療機器の安全性、効果、信頼性を評価するための指標やマトリクスの定義と重み付けが求められています。実世界での使用におけるパフォーマンス評価の時間枠についても意見を求めています。

**評価方法とインフラ**  
AI医療機器の展開後のパフォーマンスを監視するためのツールや方法論についての情報を収集しています。人間の専門家によるレビューと自動監視のバランスを取る方法、及びそれぞれの利点と欠点についても意見を求めています。

**データソースと品質管理**  
パフォーマンス評価に使用するデータソース（電子健康記録、デバイスログなど）や、データの質、完全性、相互運用性の課題への対処方法についての意見を求めています。臨床結果やユーザーフィードバックをモデル更新に組み込む効果的な方法についても関心があります。

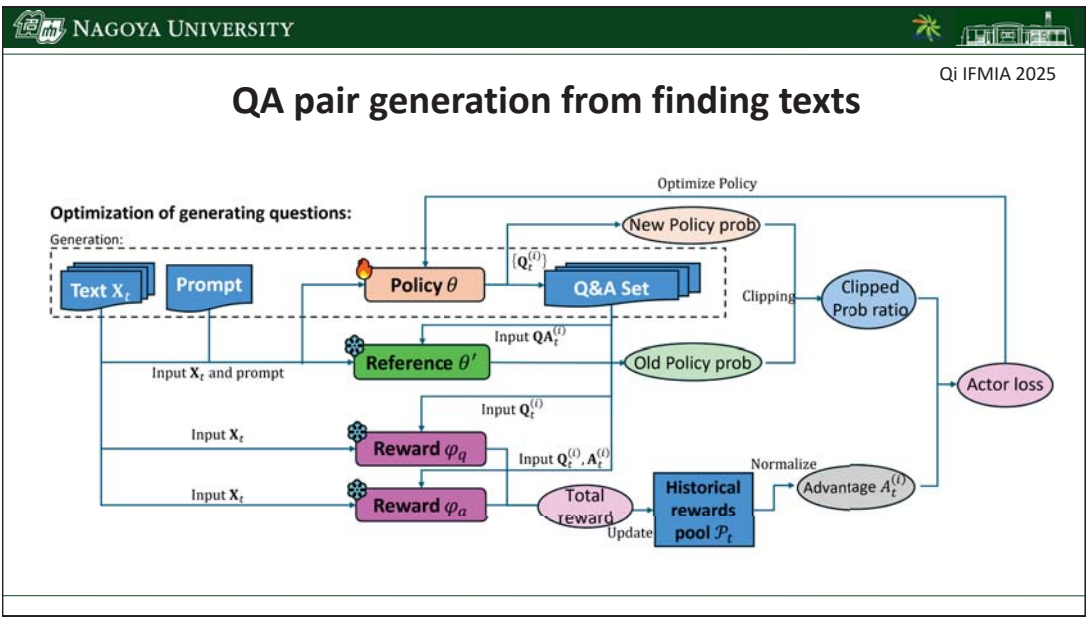
**モニタリングのトリガーと対応策**  
追加評価が必要となるトリガーや、実世界でのパフォーマンス劣化に対する定義と対応方法についての意見を求めています。

**人間とAIの相互作用**  
臨床使用パターンやユーザーの相互作用がAI医療機器のパフォーマンスに与える影響についての観察結果を求めています。安全で効果的な使用を維持するためのデザイン機能やユーザートレーニングの戦略についても意見を求めています。

**追加の考慮事項とベストプラクティス**  
実世界での検証システムの開発と実施における他の考慮事項やベストプラクティスについての意見を求めています。実施上の障壁や患者のプライバシー保護に関するアプローチについても関心があります。

<https://www.regulations.gov/document/FDA-2025-N-4203-0001>

73



**NAGOYA UNIVERSITY**

## Generated QA pairs and scores

Qi IFMIA 2025

finding1: 腹部造影CT: 前日CTと比較します。膵尾部仮性嚢胞に対してEUS-CD後、外傷チューブ留置後、前回認めた胃と脾の間の液貯留は消失しています。その他、明らかなcavityを認めません。膵頭部癌に対してGnP併用RT後、膵頭部の原発巣のサイズは著変ありません。総胆管は径約12mmで前回より拡張、腫瘍の総胆管浸潤によるものと思われる。SMA周囲の軟部影は大差ありません。主膵管拡張や膵腫大は前回より改善、急性膵炎の経過と思われず。肝・n.p.胆: 胆石あり。脾: n.p.副腎: n.p.腎: 左腎に小嚢胞あり。腹水なし。有意なサイズのリンパ節腫大なし。

Abdominal contrast-enhanced CT: Compared with the previous day's CT. After EUS-CD for a pancreatic tail pseudocyst. After placement of an external drainage tube. The fluid retention between the stomach and spleen seen previously has disappeared. No other obvious cavity is observed. After GnP with RT for pancreatic head cancer. The size of the primary lesion in the pancreatic head shows no remarkable change. The common bile duct measures about 12 mm in diameter and is more dilated than previously, thought to be due to tumor invasion of the common bile duct. The soft tissue shadow around the SMA shows no significant change. Dilatation of the main pancreatic duct and pancreatic swelling have improved compared to the previous study, considered to reflect the course of acute pancreatitis. Liver: n.p. Gallbladder: gallstones present. Spleen: n.p. Adrenal glands: n.p. Kidneys: a small cyst in the left kidney. No ascites. No lymph node enlargement of significant size.

Responses:

Q: 膵臓にどのような病変がみられましたか? score: 8.5625  
A: 膵尾部に仮性嚢胞が認められています。 score: 0.51171875

Q: 胃と脾の間に液貯留があったのですが、その状態は現在どうなっていますか? score: 2.09375  
A: 現在は消失しています。 score: 2.25

Q: 総胆管の直径はどれくらいですか? score: 15.0625  
A: 約12mmで前回よりも拡張していることが分かります。 score: 5.625

Q: 腹膜やお腹の中に異常はみられますか? score: 10.0625  
A: 腹水はなく、有意なサイズのリンパ節腫大も認められていません。 score: -5.125

Q: 膵頭部癌とは何を指すのでしょうか? score: 2.859375  
A: 膵頭部に存在する癌です。 score: 0.98828125

Q: 主膵管や膵臓自体の大きさはいかがでしょうか? score: 2.1875  
A: 前回より改善しており、急性膵炎の経過であると考えられます。 score: 0.349609375

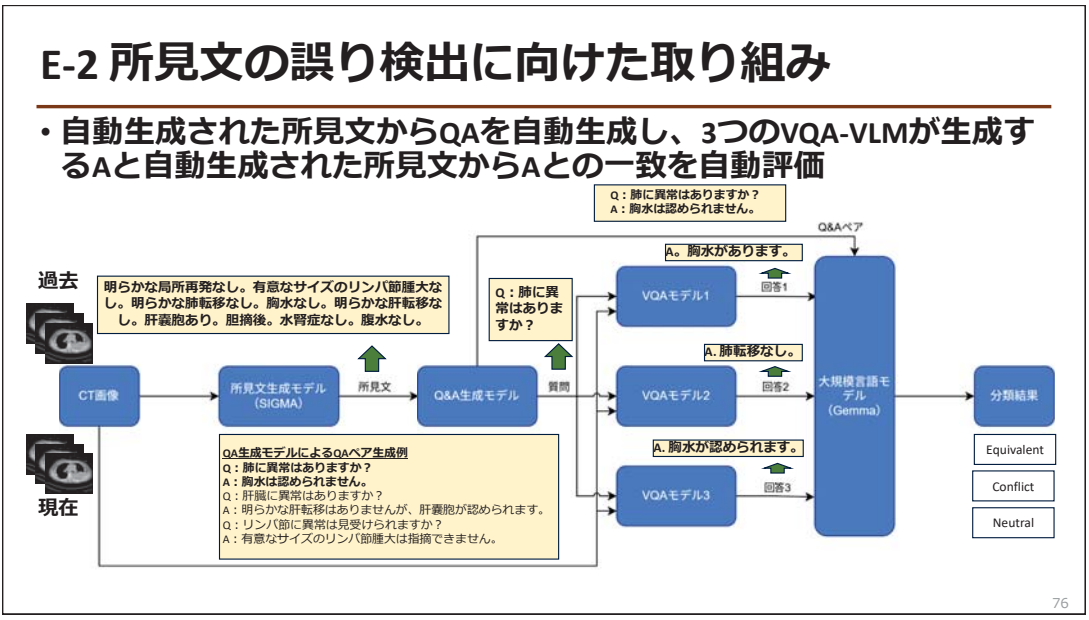
Q: There was fluid retention between the stomach and spleen; what is its status now?  
A: It has now disappeared.

Q: What is the diameter of the common bile duct?  
A: About 12 mm, and it is more dilated than before.

Q: Were there any abnormalities in the peritoneum or abdomen?  
A: There is no ascites, and no enlarged lymph nodes of significant size are observed.

Q: What does pancreatic head cancer refer to?  
A: It refers to cancer located in the pancreatic head.

Q: How are the main pancreatic duct and the pancreas itself in terms of size?  
A: They have improved compared to the previous study, considered to reflect the course of acute pancreatitis.



## 誤り検出の例

### <医師の所見文>

腹腔内に術後変化を疑うairや脂肪織濃度上昇あり。右側胸部～腹部にかけて皮下気腫あり。術後変化を疑います。肝転移なし。肝嚢胞あり。胆、脾、脾、副腎、腎に特記所見なし。肝周囲に少量腹水あり。有意な大きさのリンパ節腫大なし。両肺背側に無気肺あり。含気のある肺に明らかな活動性肺炎や転移は認めません。両側ごく少量の胸水あり。

### <生成された所見文>

明らかな局所再発は指摘できません。有意なサイズのリンパ節腫大は指摘できません。明らかな肝転移は指摘できません。肝嚢胞あり。胆嚢結石あり。胆嚢腺筋腫症(s/o)。水腎症なし。腹水なし。明らかな肺転移は指摘できません。胸水なし。

Q&Aペア	VQAの回答	LLMによる分類	実際の判定
Q: 胆嚢に異常はありますか?	VQA1 特に問題はないようです。	Conflict	Conflict
A: 胆嚢結石と胆嚢腺筋腫症が認められます。	VQA2 特に問題が指摘されていません。	Conflict	
	VQA3 著変が認められません。	Conflict	

[s/o] は「suspicion of」の略記。日本語で「～の疑い」の意

77 77

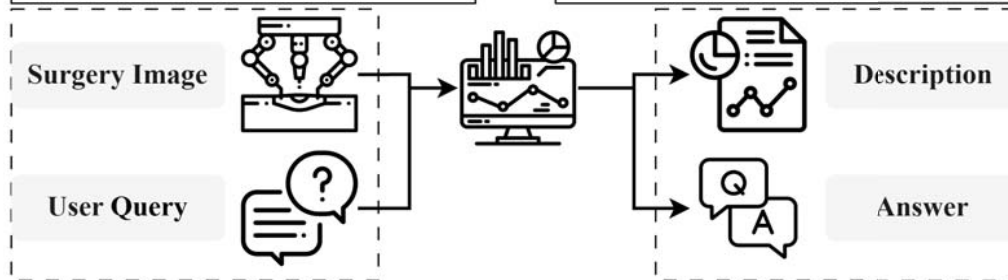
## 腹腔鏡映像に対するVQA

**Input: Image + Template Question**

eg. Please describe this image.

**Output: Description / Answer**


eg. This image shows a scene XXX.



C. Wang, SPIE Medical Imaging 2026

85

## 腹腔鏡映像に対するVQA

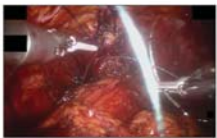


Please describe this image.

This image shows a scene from an endoscopic surgical procedure. There are 3 instruments in this picture. The Prograsp Forceps is holdsttilling. The Bipolar Forceps is holdpulling. The Monopolar Curved Scissors is traveling.

How many instruments are there in this image?

3




Locate all visible surgical instruments in this image and return their bounding boxes.

[[{"bbox\_2d": [918, 340, 1257, 566], "label": "Large Needle Driver"}, {"bbox\_2d": [0, 50, 559, 366], "label": "Large Needle Driver"}]]

Tell me the instrument's number.

There are two Large Needle Drivers in this image.



Locate the Monopolar Curved Scissors in this image and return its bounding box.

["bbox\_2d": [709, 318, 1259, 563], "label": "Monopolar Curved Scissors"]

Locate the Monopolar Curved Scissors in this image and return its bounding box.

["bbox\_2d": [528, 699, 636, 783], "label": "Suction Instrument"]

88

## Vision Language Action - 画像と言語からロボット動作コマンド

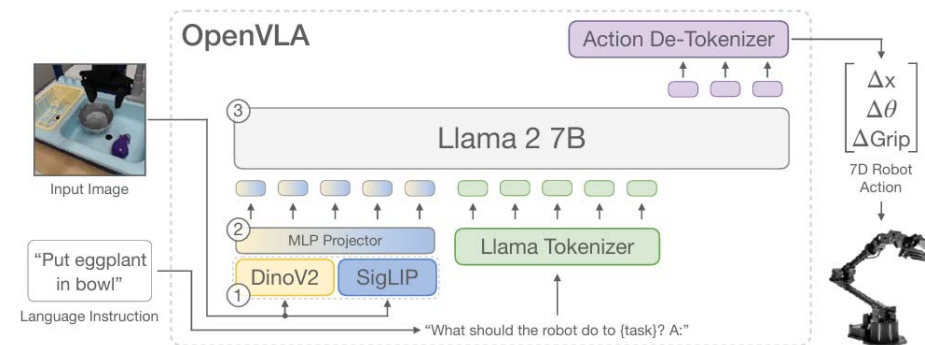


Figure 2: **OpenVLA model architecture.** Given an image observation and a language instruction, the model predicts 7-dimensional robot control actions. The architecture consists of three key components: (1) a **vision encoder** that concatenates Dino V2 [25] and SigLIP [79] features, (2) a **projector** that maps visual features to the language embedding space, and (3) the **LLM backbone**, a Llama 2 7B-parameter large language model [10].

<https://arxiv.org/pdf/2406.09246>

87



## 生成AIによる医療機器開発



**構想** アイデア整理  
**開発** プログラムのラピッドプロトタイピング  
**評価** 出力結果の自動評価  
**規制対応** データ整理 規制対応用ドキュメント作成  
**上市保険** ドキュメント作成

生成AIを用いた医療機器開発は  
 生成AIを用いた医療機器の開発だけでなく  
 医療機器開発プロセスの迅速化などを目指した生成AIの利活用も含まれる



## 生成AIによるプログラムのラピッドプロトタイピング

「眼底画像（2D RGB JPEG）から年齢を回帰推定する PyTorch コードを作成してください。ResNet50（torchvision が配布している ImageNet 事前学習モデル）を使用し、最終層のみ 1 出力の回帰に置き換えてください。train.csv ならびに val.csv には image\_path, age の列があり、test.csv には image\_path のみがあります。data/images/ に画像がある前提で、学習・検証・推論まで行い、テスト画像の推定結果を pred\_test.csv に保存するフルスクリプトを作成してください。」（ChatGPTによるプロンプト修正を含む）

```

import os
from typing import Tuple
import pandas as pd
from PIL import Image
import torch
import torch.nn as nn
from torch.utils.data import Dataset, DataLoader
from torchvision import models, transforms

# Dataset 定義
class FundusAgeTrainValDataset(Dataset):
    """眼底画像 + 年齢(回帰)用 Dataset"""
    def __init__(self, csv_path: str, img_root: str, transform=None):
        self.csv = pd.read_csv(csv_path)
        self.img_root = img_root
        self.transform = transform

    def __len__(self):
        return len(self.csv)

    def __getitem__(self, index):
        image_path = self.csv.iloc[index].image_path
        image = Image.open(os.path.join(self.img_root, image_path))
        if self.transform:
            image = self.transform(image)
        age = self.csv.iloc[index].age
        return image, age

# Train / Eval
def train_one_epoch(model: nn.Module, dataloader: DataLoader, criterion, optimizer, device: torch.device, epoch: int) -> Tuple[float, float]:
    model.train()
    running_loss = 0.0
    running_mse = 0.0
    total = 0
    for data_loader_iter in dataloader:
        images, ages = data_loader_iter
        images, ages = images.to(device), ages.to(device)
        outputs = model(images)
        loss = criterion(outputs, ages)
        optimizer.zero_grad()
        loss.backward()
        optimizer.step()
        running_loss += loss.item()
        running_mse += ((outputs - ages).pow(2)).mean().item()
        total += 1
    return running_loss / total, running_mse / total

if "image_path" not in self.csv.columns or "age" not in self.csv.columns:
    raise ValueError("CSVには 'image_path' と 'age' 列が必要です。")
    
```

ChatGPTによるコード生成例

迅速にプロトタイピングが可能であり医療機器開発の加速に貢献



## FDA Total Product Lifecycle Considerations for Generative AI Enabled DevicesにおけるAI Lifecycle

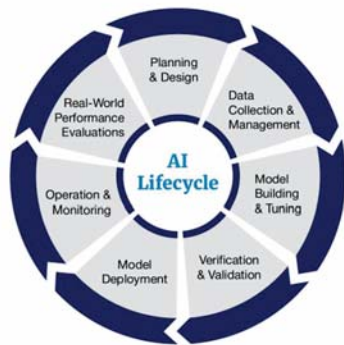
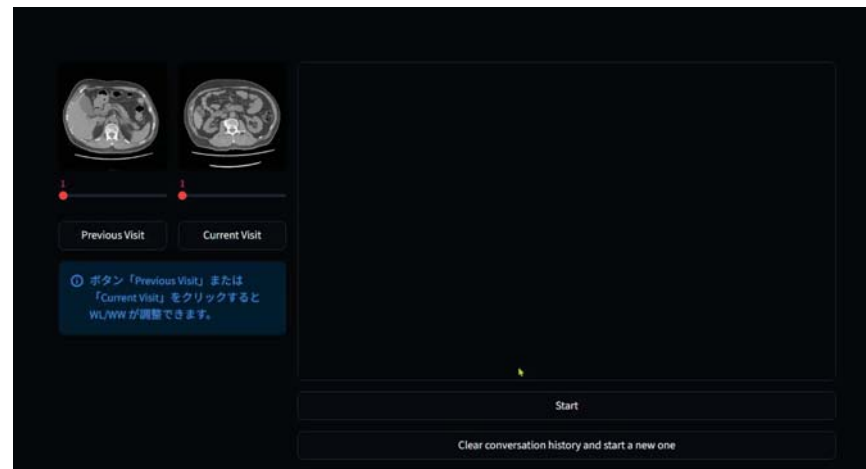


Figure 1. AI Lifecycle



## AI vs AI (Qを作りAを返しQを作りAを返す)





# 1960年代の対話Eliza (人工無能とも呼ばれる)

```

Linux taka 5.10.0-36-686-pae #1 SMP Debian 5.10.244-1 (2025-09-29) 1686
I updates could not be installed automatically. For more details,
see /var/log/unattended-upgrades/unattended-upgrades.log

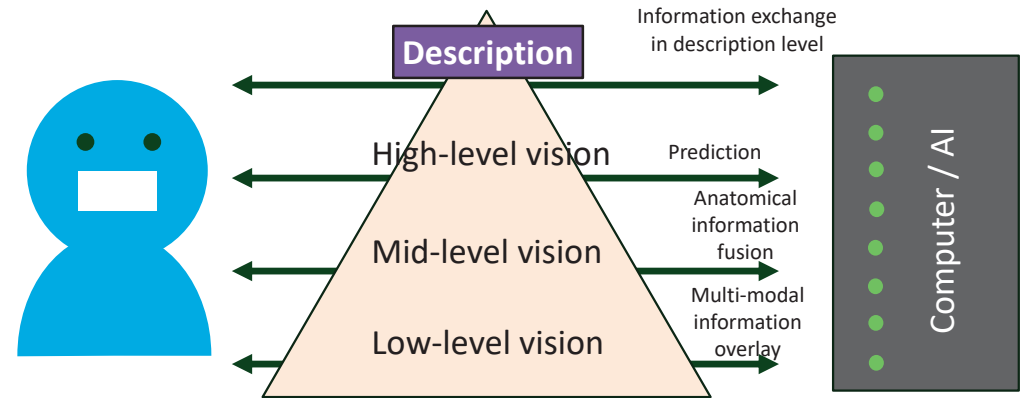
The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
You have mail.
Last login: Fri Feb 20 17:40:00 2026 from 10.0.0.1
mori@taka:~$
mori@taka:~$
mori@taka:~$ emacs
mori@taka:~$ emacs
mori@taka:~$

```



# 言語を介したAIとの対話 –画像特徴言語化の時代-



## 生成AI活用医療機器は 承認審査をどう変えるか！？

独立行政法人 医薬品医療機器総合機構(PMDA)  
プログラム医療機器審査部  
加藤健太郎

本発表は、発表者の個人的見解に基づくものであり、  
独立行政法人医薬品医療機器総合機構の公式見解を示すものではありません。



## 本日のテーマ

# 生成AIを活用した医療機器の審査は 何が特殊か？



## おことわり

### 医療機器プログラムと汎用AIの違いについて

#### 医療機器プログラム

- ・医療機器としての目的性(疾病の診断、治療等に寄与するなど)を有しており、かつ、意図したとおりに機能しない場合に患者(又は使用者)の生命及び健康に影響を与えるおそれがあるプログラム(ソフトウェア機能)。
- ・提供するためには、医薬品医療機器等法に基づく製造販売承認や認証を事前に取得することが必要。

#### その他(汎用AIなど)のプログラム

- ・疾病の診断や予防、治療の目的を標榜せずに、提供されるプログラム。医療用途以外の一般的な目的で提供される、汎用AIなどが該当。

➤疾病の診断や予防、治療に用いる医療機器プログラムは、医療機器として、有効性や安全性が確認された上で提供されている。

➤一方、汎用AIなどのその他のプログラムは医療機器として承認・認証されたものではなく、疾病の診断や予防、治療の目的を標榜して提供することはできない。また、健康状態や疾病に関する質問をした場合の回答内容を含めたその性能は、医薬品医療機器等法に基づき、その妥当性が確認されたものではない。

[https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000179749\\_00004.html](https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000179749_00004.html)

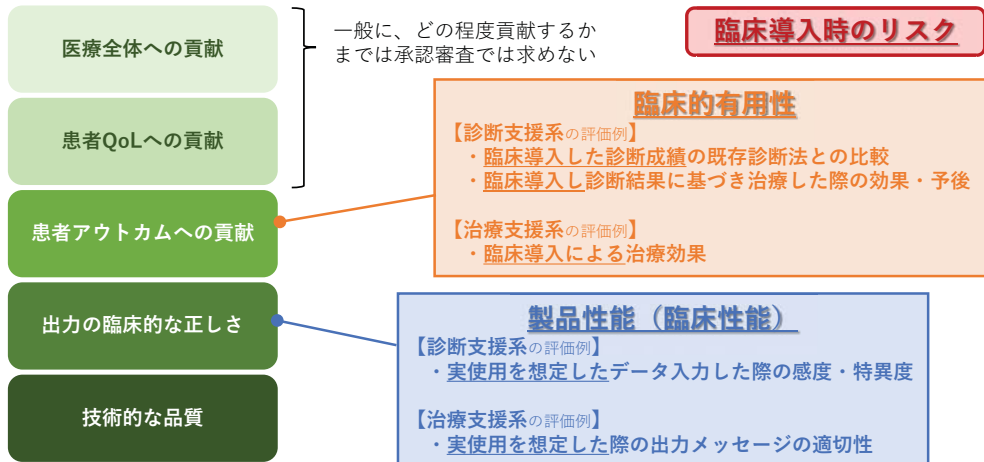
## おことわり

生成AIの活用範囲は多岐にわたり、既に一部利活用も進んでいます  
何が規制されるべき製品であるか(医療機器該当性)は  
別途慎重な検討が必要です

また、PMDAでは生成AIを活用した医療機器の経験はほとんどありません

本講演は、“医療機器に該当した場合”について  
“従来の審査経験に基づく”  
審査上どのような特殊性があるかを考察するものです。

## 審査で考えていること



“臨床導入を想定”するケースが

現実可能な範囲で想定できれば

従来の審査方針の応用は比較的容易

## 審査上課題となるような生成AIとは？

使用者からの自然文や音声による様々な入力に応じて  
特定のタスクに依らない多様な回答や創造的なコンテンツを  
生成する生成AIを活用したもの

- 多様な自然文・音声入力、使用者との相互的なやり取りが可能
- 特定のタスクに依らない多様かつ創造的なコンテンツの生成
- 最新の知識や入力情報をリアルタイムで取り入れた出力生成が可能

入出力の多様性が膨大

使用者の使い方の  
影響が大きい

製品状態の固定が困難

## 生成AIの特性と評価上の課題

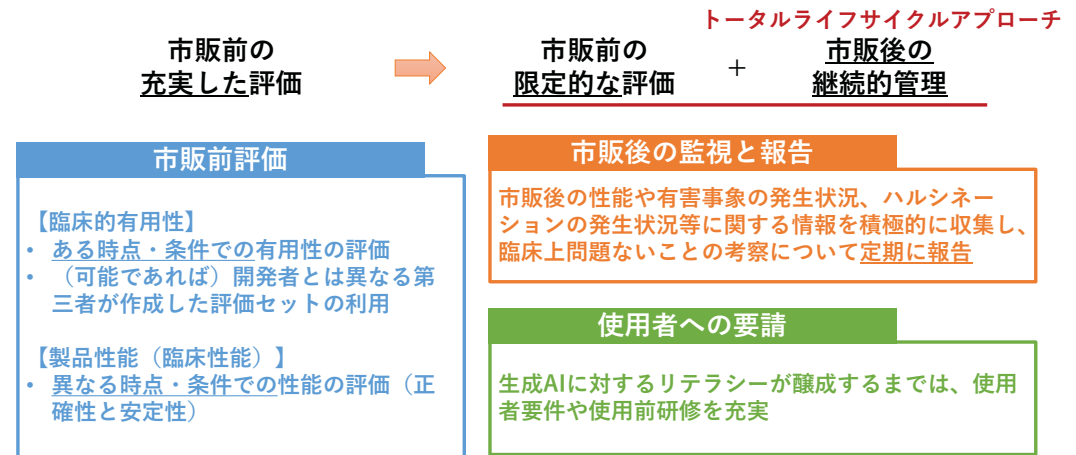
	特性	主な課題
運用経験に基づく学習性	パラメータの更新を伴うことなく製品の振舞いを調整することができる性質	<ul style="list-style-type: none"> <li>どの時点の製品の評価を行うべきか</li> <li>調整され続ける製品の有効性・安全性をどのように管理するか</li> </ul>
動的参照性	Web情報や外部データベースなど動的に変化する情報源をリアルタイムに参照して出力を生成する性質	<ul style="list-style-type: none"> <li>参照する情報が変化するために、同じ入力をしていても評価時点により出力が変化する</li> <li>Web上にある情報を制限なく活用する場合には、誤情報等や偏った情報を取り込み、不正確な回答を生成する可能性がある</li> </ul>
データバイアス	特に極めて大規模なデータを用いて構築されたモデルの場合、多量・多様な非標準化データを用いて学習をしているため、内在するバイアスの特定や制御がより難しくなる	<ul style="list-style-type: none"> <li>バイアスがどのように表在化するか確認することが困難</li> </ul>
出力の非決定性	同一入力に対して毎回異なる出力が得られるという性質	<ul style="list-style-type: none"> <li>ある入力に対して1回の出力だけを評価しても、その製品全体の出力を評価できていない</li> </ul>
生成コンテンツに対する要解釈性	生成コンテンツの解釈（臨床への反映）が使用者に強く依存する	<ul style="list-style-type: none"> <li>自然文などの生成コンテンツにおける意味内容を客観性をもってどのように評価するか</li> </ul>

## 使用上の課題

- 出力の正確性や妥当性を、使用者が判断できないことがある。また、ハルシネーションに気づけない場合がある。
- 人よりAIを信じてしまい、助言を過大評価するリスク(自動化バイアス)がある。
- 自由入力のため、適応外・禁忌・根拠の弱い入力が行われる可能性がある。
- 意図せず適応外のコンテンツ(回答)を生成してしまう可能性がある。



## これを踏まえて何ができるか



## その他の事項

特に、以下のような場合は、関連学会等のステークホルダーとの連携も重要になる。

- 従来の診療や治療のフローを変えるもの。
- 生成AIの出力の解釈や妥当性判定が明らかに困難であるもの。
- 自動化バイアスの促進により、医療者と患者のコミュニケーションに影響を与える可能性があるもの。



## 最後に

本内容は、**審査経験に基づくただのアイデア**です

より合理的な方法やアイデアは大歓迎です！

よりよい医療のため、よりよい開発環境実現のため  
是非お知恵を貸してください！！



# PMDA

健やかに生きる世界を、ともに

ご清聴ありがとうございました！  
ご質問等ありましたら[kato-kentaro@pmda.go.jp](mailto:kato-kentaro@pmda.go.jp)まで！



Copyright © Pharmaceuticals and Medical Devices Agency, All Rights Reserved.