



# 厚労科研AI開発1年目 の総括と今後

東京科学大学 石田 茂樹



石田茂樹

東京科学大学  
情報理工学院 情報工学系  
横田研究室 修士2年

研究テーマ:

大規模言語モデルの医療応用  
Mechanistic Interpretability

1

医療用LLM

2

構造化用LLM開発

3

まとめ

1

LLMの医療応用

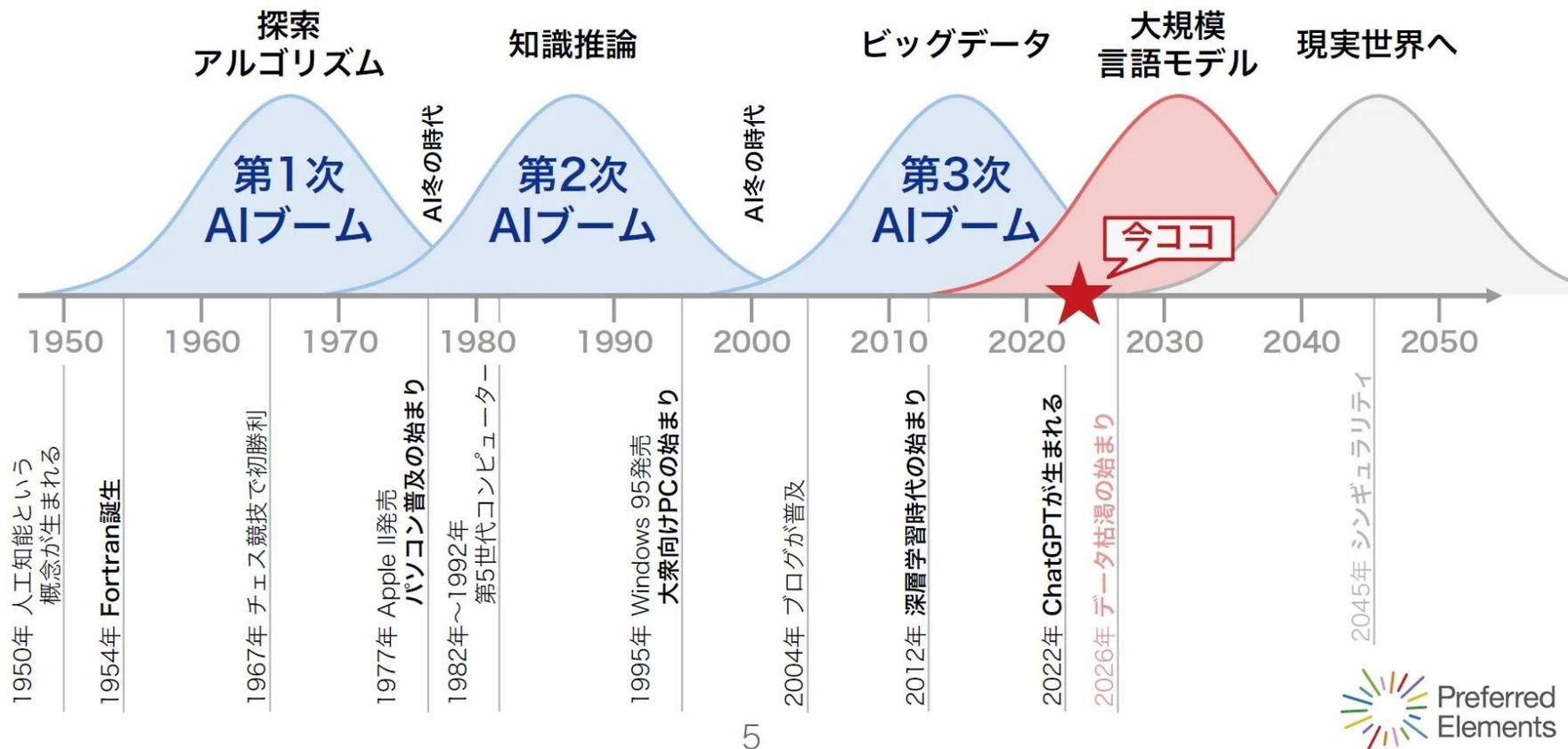
2

構造化用LLM開発

3

まとめ

## AIブームの歴史: 第4次AIブームは始まったばかり





<https://gentosha-go.com/articles/-/66189>

「医療」に特化したオープンソースの大規模言語モデル「Meditron」が登場



Metaの大規模言語モデル(LLM)である「Llama 2」をベースに医学書でトレーニングした、医療専門のLLMスイート「Meditron」が登場しました。登場直後であるため正規採用はまだ推奨されていないものの、医療分野に関する能力がGPT-3.5を上回っていることが報告されています。

<https://gigazine.net/news/20231203-meditron-medical-pretraining-llm/>

## Googleの医療AI「AMIE」が優れた鑑別診断をアシスト

執筆者 R.A. 2025年4月30日



<https://aitimes.media/2025/04/30/15285/>

## ChatGPTが医師国家試験「合格」も、診療利用に不向きな理由

有料記事

枝松佑樹 2023年6月15日 17時30分

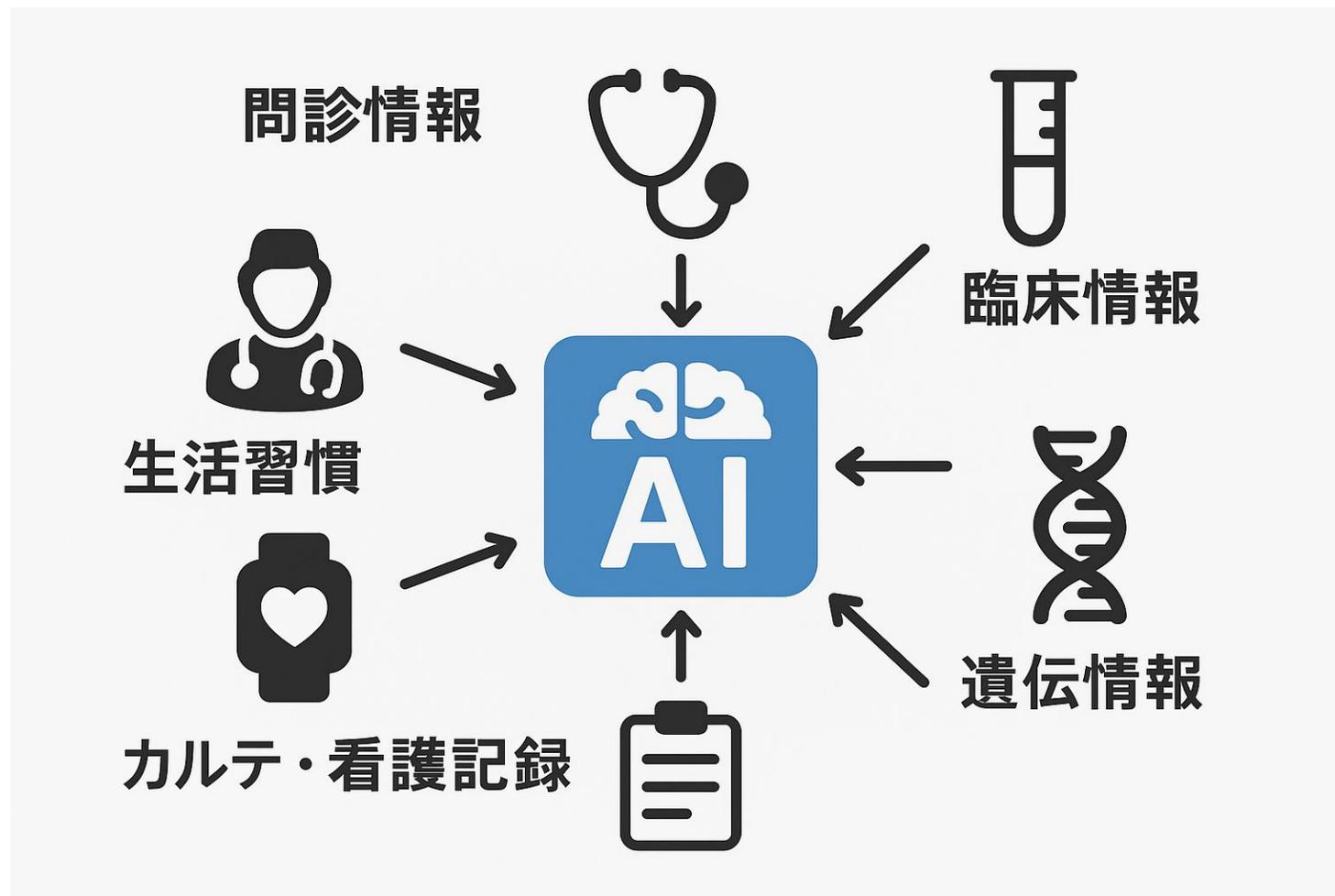
<https://www.asahi.com/articles/ASR6H3Q13R6GUTFL00H.html>

- 国内ではPreferred Networksが医療ドメインの独自コーパスを用いた finetuning を行い、医療特化の大規模言語モデル Llama3-Preferred-MedSwallow-70B を開発

モデル名	2018	2019	2020	2021	2022	平均
Llama3-Preferred-MedSwallow-70B	407	390	391	393	395	395.2
GPT-4	382	385	387	398	392	388.8
Llama-3-Swallow-70B-v0.1	353	347	353	345	345	348.6
Meta-Llama-3-70B	353	340	348	314	318	334.6
Qwen2-72B	320	325	325	326	360	331.2
c4ai-command-r-plus	321	303	320	302	335	316.2
gemma-2-27b	337	298	327	296	322	316
Swallow-70b-NVE-hf	283	280	300	295	300	291.6
Swallow-MX-8x7b-NVE-v0.1	269	285	290	277	290	282.2
ChatGPT	266	250	266	297	287	273.2

<https://tech.preferred.jp/ja/blog/llama3-preferred-medswallow-70b/>

LLMは医療現場でどのようなことが期待されているのか



1

医療用LLM

2

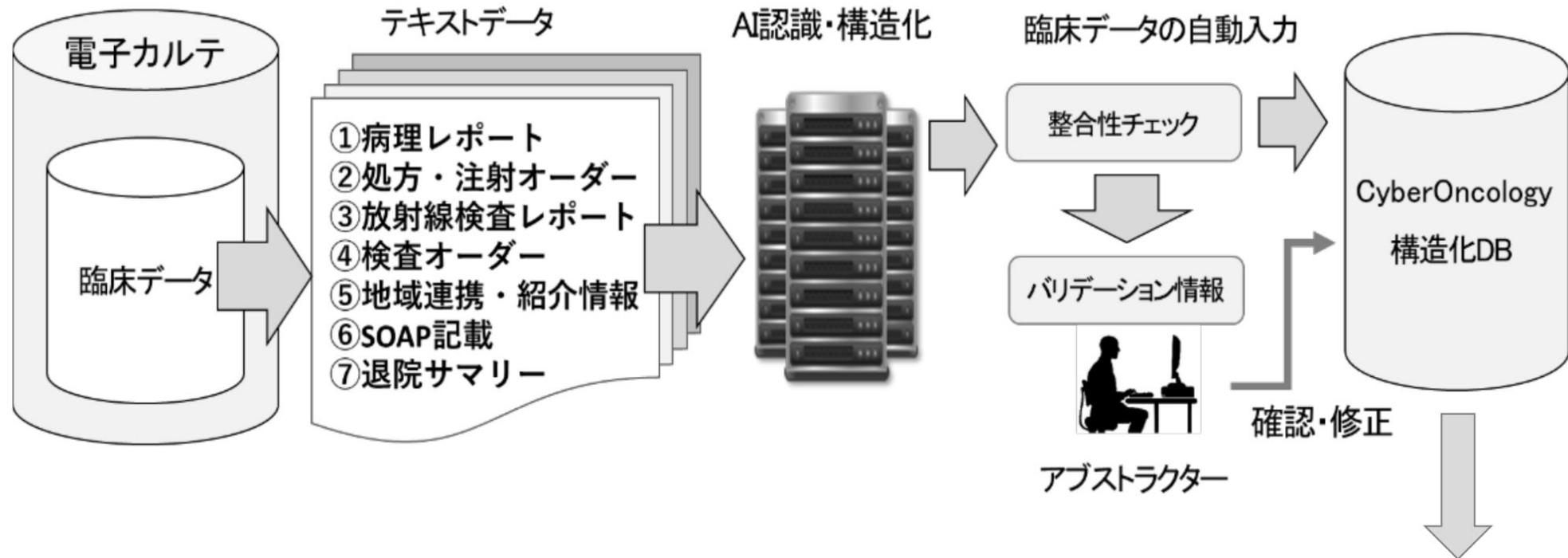
**構造化用LLM開発(厚労科研)**

3

まとめ

# LLMの医療応用事例

- 電子カルテデータをLLMに学習させることで医療情報を抽出&構造化



厚労科研:大規模言語モデル(LLM:Large Language Model)を活用した医薬品等の有効性・安全性評価のためのアウトカム抽出の方法論の確立に向けた研究(24AC0401)

- C-CAT
- 退院サマリー、紹介状 (SIP3研究)
- 各学会レジストリ (認定医、専門医、指導医)

## 【構造化評価文: 電子カルテの経過記録例】

2019年12月初旬に胸痛、背部痛を自覚。近医を受信し左肺門部腫瘍を指摘され、1月7日に当院を紹介受診。同日に胸腔穿刺を施行し、得られた胸水から肺腺癌と診断

T4N1M1c, StageIVC, BRA, PUL, LYM, PLE, OSS, EGFR(L858R+), KRAS-, BRAF(V600E)-, ROS1-, PDL-1 < 0%, ALK-

1/31~: C1-7頸椎転移に対し緩和的放射線治療(30Gy/10fr)施行

2/6~: 1stline Osimertinib 80mg/dayを開始(28日分)。

11/4: CTにて左肺門部腫瘍、左肺尖結節、肝転移、骨転移増大を認めPDと判断。

11/19~2/10: 2nd line CBDCA/PTX/Bev/Atezoを4course施行

2021/3/2: 効果判定にて原疾患の増悪を認めPDと判断。

3/16~4/13: 3rd line DOC+RAM 2course施行

5/18: 効果判定にて両肺の小結節は増加・増大と両側胸水の増加がありPDと判断

5/20: 胸水コントロール目的に入院。

5/21: 左癌性に対して胸水左胸腔ドレーン挿入

5/24: 左胸膜癒着術(ユニタルク4g)を施行。

5/27~: 4th line EGFR-TKI rechallenge(Afatinib 20mg/day)開始

LLM

## 【LLMによるJSON形式出力例】

```
.....  
{  
  "date": "2020年1月7日",  
  "event": "胸腔穿刺を施行し、得られた胸水から肺腺癌と診断",  
  "diagnosis": "肺腺癌",  
  "treatment": "胸腔穿刺",  
  "TNM_classification": "T4N1M1c",  
  "stage": "StageIVC",  
  "metastasis": "BRA, PUL, LYM, PLE, OSS",  
  "specimen": "胸水",  
  "gene_mutation": "EGFR(L858R+), KRAS-, BRAF(V600E)-, ROS1-,  
PDL-1<0%, ALK-"  
},  
{  
  "date": "2020年1月31日~",  
  "event": "C1-7頸椎転移に対し緩和的放射線治療(30Gy/10fr)施行",  
  "diagnosis": null,  
  "treatment": "緩和的放射線治療",  
  "TNM_classification": null,  
  "stage": null,  
  "metastasis": "C1-7頸椎転移",  
  "specimen": null,  
  "gene_mutation": null  
},  
.....
```

図1. LLMによる電子カルテ情報の構造化例

## 電子カルテの各臨床項目の内容が含まれているかを評価

事象	経過記録	抽出臨床項目	項目数
1	2019年12月初旬に胸痛、背部痛を自覚。近医を受信し左肺門部腫瘍を指摘され	日時、症状	2
2	1月7日に当院を紹介受診。同日に胸腔穿刺を施行し、得られた胸水から肺腺癌と診断 T4N1M1c StageIVC, BRA, PUL, LYM, PLE, OSS, EGFR(L858R+), KRAS-, BRAF(V600E)-, ROS1-, PDL-1<0%, ALK-	診療日、検体キー、検体名、診断キー、診断名、TNMキー、TNM分類、Stageキー、Stage、指摘キー、誤字やStageの間違いの指摘、転移キー、転移値(5つ)、バイオマーカーキー(6つ)、バイオマーカー値(6つ)	30
3	1/31～: C1-7頸椎転移に対し緩和的放射線治療(30Gy/10fr)施行	開始日、治療キー、治療内容、部位キー、部位名、線量キー、線量値	7
4	2/6～: 1stline Osimertinib 80mg/dayを開始(28日分)。	開始日、ラインキー、治療ライン、治療キー、治療タイプ、薬剤キー、薬剤名、投与キー、投与量、期間キー、投与期間	11
5	11/4: CTにて左肺門部腫瘍、左肺尖結節、肝転移、骨転移増大を認めPDと判断。	診療日、判定手段、検査内容、判定内容	4
6	11/19～2/10: 2nd line CBDCA/PTX/Bev/Atezoを4course施行	開始日、終了日、ラインキー、治療ライン、治療キー、治療タイプ、薬剤キー、薬剤名、投与キー、投与量	10
7	2021/3/2: 効果判定にて原疾患の増悪を認めPDと判断。	診療日、判定手段、検査内容、判定内容	4
8	3/16～4/13: 3rd line DOC+RAM 2course施行	開始日、終了日、ラインキー、治療ライン、治療キー、治療タイプ、薬剤キー、薬剤名、投与キー、投与量	10
9	5/18: 効果判定にて両肺の小結節は増加・増大と両測胸水の増加がありPDと判断	診療日、判定手段、検査内容、判定内容	4
10	5/20: 胸水コントロール目的に入院。	診療日、治療キー、治療内容	3
11	5/21: 左癌性に対して胸水左胸腔ドレーン挿入	診療日、治療キー、治療内容	3
12	5/24: 左胸膜癒着術(ユニタルク4g)を施行	診療日、治療キー、治療内容	3
13	5/27～: 4th line EGFR-TKI rechallenge(Afatinib 20mg/day)開始	開始日、ラインキー、治療ライン、治療キー、治療タイプ、薬剤キー、薬剤名、投与キー、投与量	9
抽出項目数の合計			100

## 電子カルテの各臨床項目の内容が含まれているかを評価

日時、検体、診断、TNM、Stage、転移(5つ)、バイオマーカー(6つ)

事象	経過記録	抽出臨床項目	項目数
1	2019年12月初旬に胸痛、背部痛を自覚。近医を受信し左肺門部腫瘍を指摘され	日時、症状	2
2	1月7日に当院を紹介受診。同日に胸腔穿刺を施行し、得られた胸水から肺腺癌と診断 T4N1M1c StageIVC, BRA, PUL, LYM, PLE, OSS, EGFR(L858R+), KRAS-, BRAF(V600E)-, ROS1-, PDL-1<0%, ALK-	診療日、検体キー、検体名、診断キー、診断名、TNMキー、TNM分類、Stageキー、Stage、指摘キー、誤字やStageの間違いの指摘、転移キー、転移値(5つ)、バイオマーカーキー(6つ)、バイオマーカー値(6つ)	30
3	1/31～: C1-7頸椎転移に対し緩和的放射線治療(30Gy/10fr)施行	開始日、治療キー、治療内容、部位キー、部位名、線量キー、線量値	7
4	2/6～: 1stline Osimertinib 80mg/dayを開始(28日分)。	開始日、ラインキー、治療ライン、治療キー、治療タイプ、薬剤キー、薬剤名、投与キー、投与量、期間キー、投与期間	11
5	11/4: CTにて左肺門部腫瘍、左肺尖結節、肝転移、骨転移増大を認めPDと判断。	診療日、判定手段、検査内容、判定内容	4
6	11/19～2/10: 2nd line CBDCA/PTX/Bev/Atezoを4course施行	開始日、終了日、ラインキー、治療ライン、治療キー、治療タイプ、薬剤キー、薬剤名、投与キー、投与量	10
7	2021/3/2: 効果判定にて原疾患の増悪を認めPDと判断。	診療日、判定手段、検査内容、判定内容	4
8	3/16～4/13: 3rd line DOC+RAM 2course施行	開始日、終了日、ラインキー、治療ライン、治療キー、治療タイプ、薬剤キー、薬剤名、投与キー、投与量	10
9	5/18: 効果判定にて両肺の小結節は増加・増大と両測胸水の増加がありPDと判断	診療日、判定手段、検査内容、判定内容	4
10	5/20: 胸水コントロール目的に入院。	診療日、治療キー、治療内容	3
11	5/21: 左癌性に対して胸水左胸腔ドレーン挿入	診療日、治療キー、治療内容	3
12	5/24: 左胸膜癒着術(ユニタルク4g)を施行	診療日、治療キー、治療内容	3
13	5/27～: 4th line EGFR-TKI rechallenge(Afatinib 20mg/day)開始	開始日、ラインキー、治療ライン、治療キー、治療タイプ、薬剤キー、薬剤名、投与キー、投与量	9
抽出項目数の合計			100

単に医療データで学習を行うだけでは、構造化データのような実際の医療現場で必要とされる事例では**不十分の可能性**

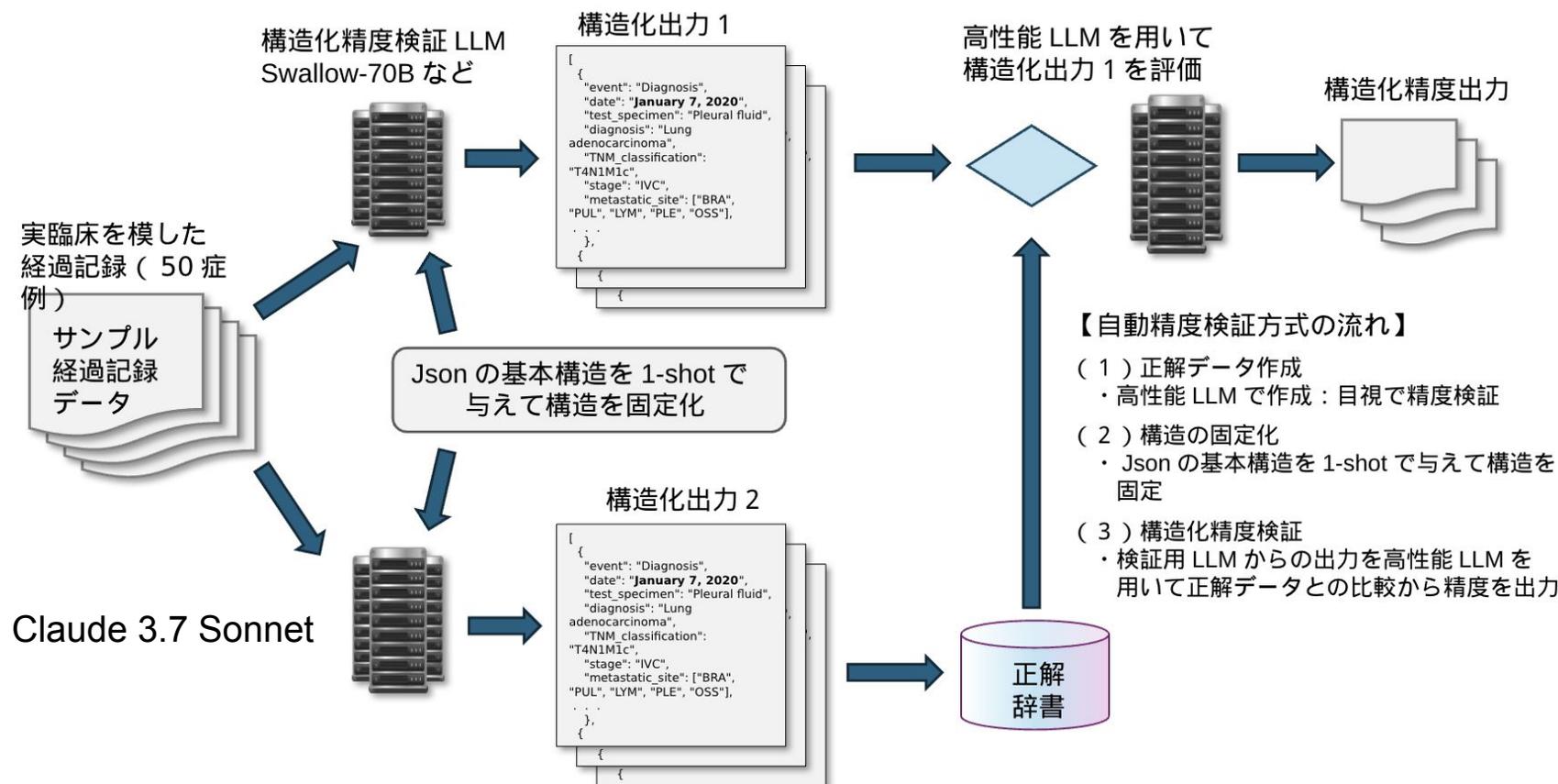


MedSwallow

事象 項番	経過記録	臨床項目	項目 数	Meta-Llama-3-70B								Llama-3-Swallow-70B								Llama3-Preferred-MedSwallow-70B							
				O2	O3	O4	O5	O6	O8	なし	O2	O3	O4	O5	O6	O8	なし	O2	O3	O4	O5	O6	O8	なし			
1	2019年12月初旬に胸痛、背部痛を自覚。近医を受信し左肺門部腫瘍を指摘され	診療日、症状	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	0	0	0	2	0	0		
2	1月7日に当院を紹介受診。同日に胸腔穿刺を施行し、得られた胸水から肺腺癌と診断 T4N1M1c StageIVC, BRA, PUL, LYM, PLE, OSS, EGFR(L858R+), KRAS-, BRAF(V600E)-, ROS1-, PDL-1<0% ALK-	診療日	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0			
		検体キー、検体名	2	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
		診断キー、診断名	2	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0	2	2		
		TNMキー、TNM分類	2	1	2	2	2	2	2	2	1	2	2	2	2	2	2	1	0	0	0	1	0	2			
		Stageキー、Stage	2	0	0	2	2	2	2	2	0	1	2	2	2	2	2	0	0	0	0	1	0	2			
		指摘キー、誤字やStage間違いの指摘	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
		転移キー、転移値(5つ)	6	6	0	6	6	6	6	6	0	4	6	6	0	0	6	0	4	3	1	5	4	6			
		バイオマーカーキー(6つ)	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6		
バイオマーカー値(6つ)	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6				
3	1/31~ :C1-7頸椎転移に対し緩和的放射線治療(30Gy/10fr)施行	開始日、治療キー、治療内容、部位キー、部位名、線量キー、線量値	7	4	7	5	7	5	7	7	0	5	5	5	7	7	7	2	0	7	4	0	0	7			
4	2/6~ :1stline Osimertinib 80mg/ dayを開始(28日分)	開始日、ラインキー、治療ライン、治療キー、治療タイプ、薬剤キー、薬剤名、投与キー、投与量、期間キー、投与期間	11	7	9	7	10	11	11	11	10	11	9	10	10	11	11	3	8	9	10	6	9	9			
5	11/4 :CTにて左肺門部腫瘍、左肺尖結節、肝転移、骨転移増大を認めPDと判断。	診療日、判定手段、検査内容、判定内容	4	1	0	3	4	4	4	4	3	3	3	3	4	4	4	0	0	2	2	3	0	4			
6	11/19~ 2/10 :2nd line CBDCA/ PTX/ Bev/ Atezoを4course施行	開始日、終了日、ラインキー、治療ライン、治療キー、治療タイプ、薬剤キー、薬剤名、投与キー、投与量	10	6	7	6	8	10	10	10	8	9	7	9	9	10	10	2	8	8	9	2	8	8			
7	2021/3/2 効果判定にて原疾患の増悪を認めPDと判断。	診療日、判定手段、検査内容、判定内容	4	3	4	4	4	4	4	4	3	4	3	3	4	4	4	2	0	2	2	3	0	4			
8	3/16~ 4/13 :3rd line DOC+RAM 2course 施行	開始日、終了日、ラインキー、治療ライン、治療キー、治療タイプ、薬剤キー、薬剤名、投与キー、投与量	10	7	7	6	8	10	10	10	8	9	7	9	10	10	10	2	8	8	9	4	8	8			
9	5/18 効果判定にて両肺の小結節は増加・増大と両側胸水の増加がありPDと判断	診療日、判定手段、検査内容、判定内容	4	3	4	4	4	4	4	4	3	3	3	3	4	4	4	2	0	2	2	3	0	4			
10	5/20 胸水コントロール目的に入院。	診療日、治療キー、治療内容	3	2	3	3	3	3	3	3	3	3	0	3	3	0	0	0	0	0	0	0	0	0			
11	5/21 左癌性に対して胸水左胸腔ドレーン挿入	診療日、治療キー、治療内容	3	2	3	3	3	3	3	3	3	3	0	3	3	3	3	0	0	0	0	0	0	0			
12	5/24 左胸膜癒着術(ユニタルク4q)を施行	診療日、治療キー、治療内容	3	2	3	3	3	3	3	3	3	3	0	3	3	3	3	0	0	0	0	0	0	0			
13	5/27~ :4th line EGFR- TKI rechallenge (Afatinib 20mg/ day)開始	開始日、ラインキー、治療ライン、治療キー、治療タイプ、薬剤キー、薬剤名、投与キー、投与量	9	7	7	5	8	9	9	9	8	9	6	8	8	9	9	4	7	7	7	7	7	7			
正解項目数の合計 (= 構造化精度%)			100	67	72	77	90	94	96	96	69	86	70	86	86	86	92	35	49	62	60	49	50	75			

- LLMが出力するjsonが評価項目を抽出できているかどうかをLLMで評価する手法を確立

※GPT-4のAPIを個人で用意することが難しかったので、最先端モデルとしてGPT-4にもひけを取らないClaude 3.7 Sonnetにより評価を行った

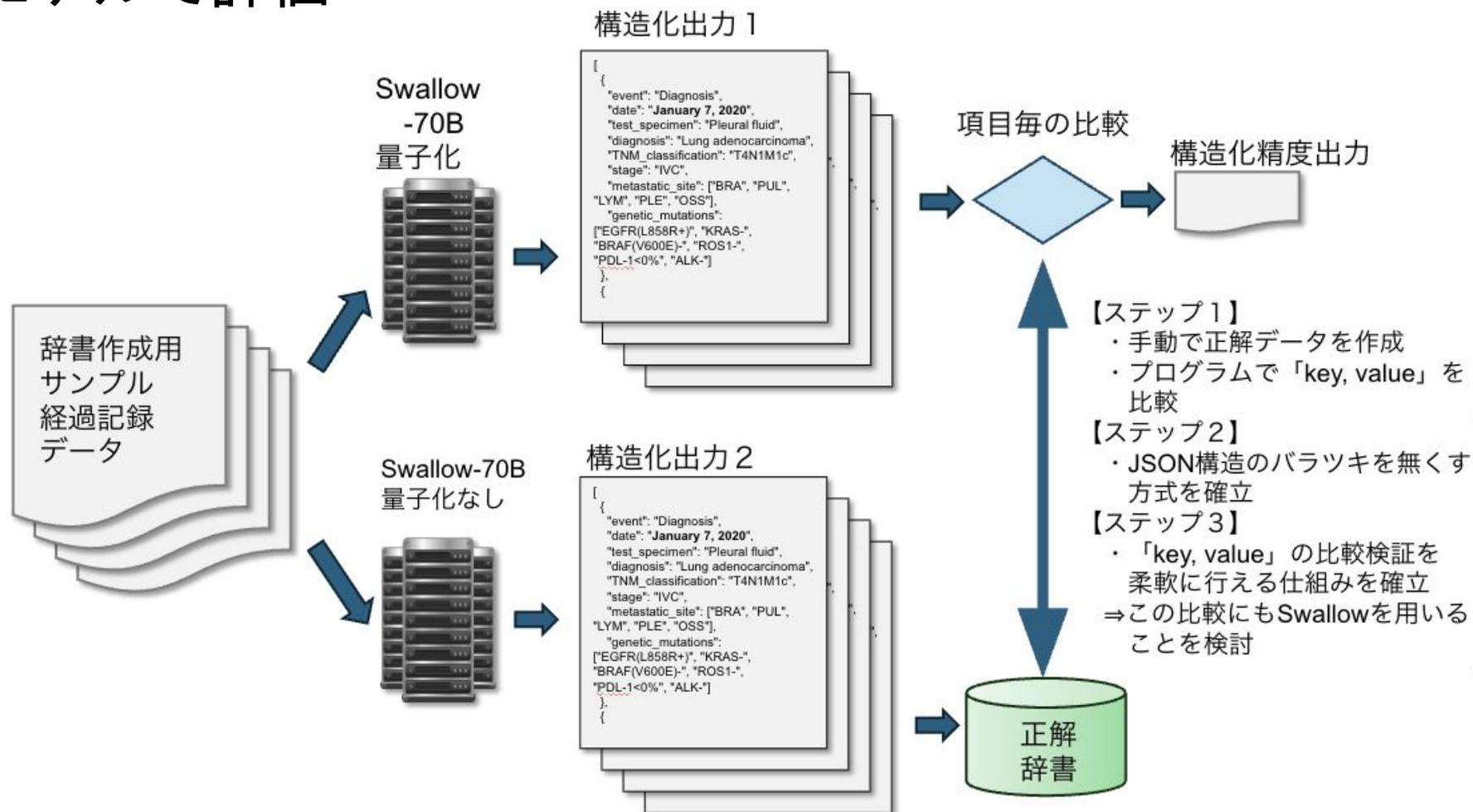


\* 高性能 LLM

当初 Llama 3.1-405B で試みたが日本語での評価が不十分のため、Anthropic-Claude-3.5-Sonnet に変更

# LLMによる評価方法を自動化する方法を確立

- 臨床現場では70Bパラメータのモデルを大量に動かすことは難しい
  - 量子化モデルを使って構造化出力
  - 量子化なしのモデルで評価



- 全臨床項目を適切に評価
- 正解データとの照合に成功
- 各事象に対して妥当な採点を実施

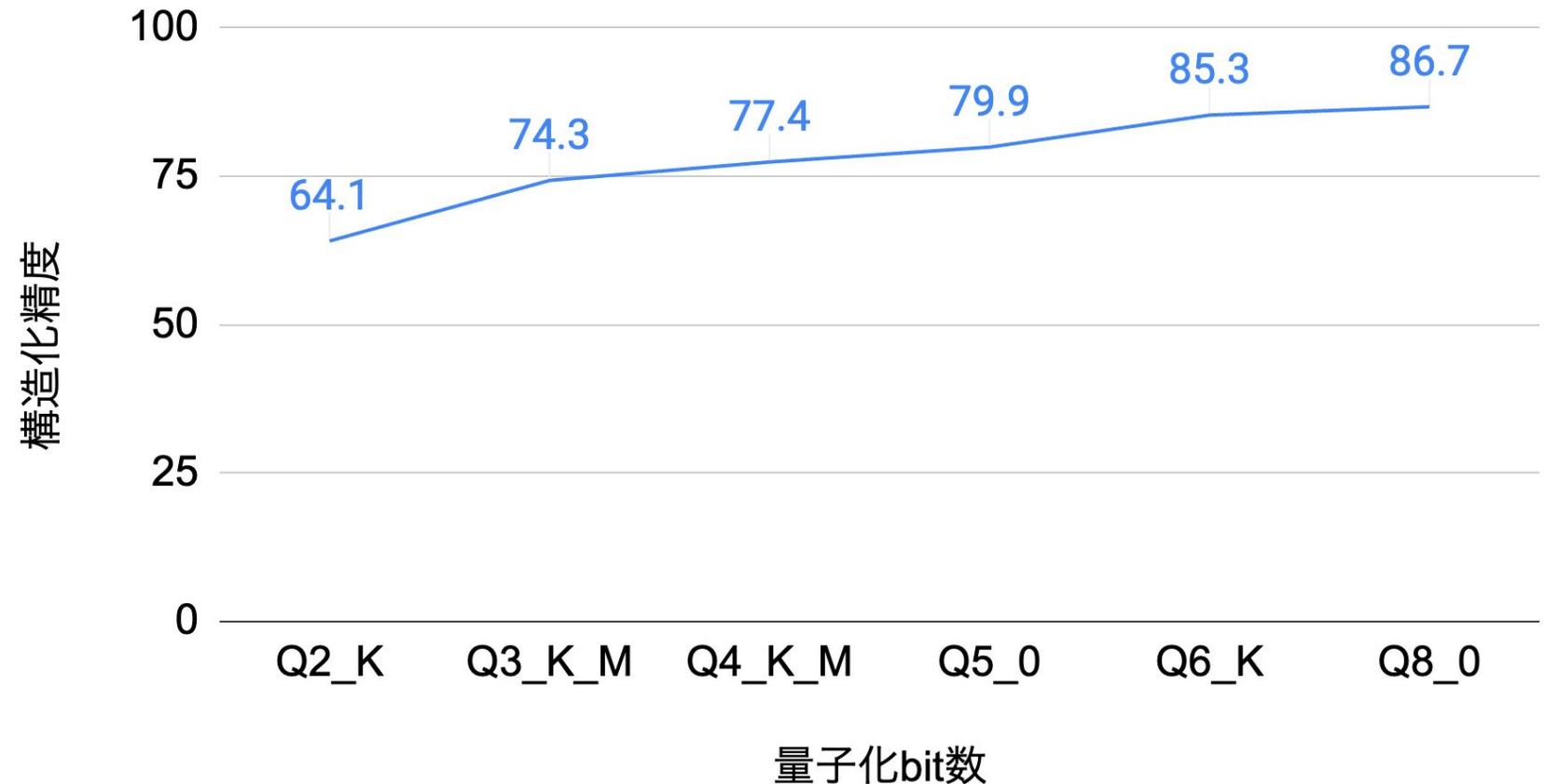
診療日、検体キー、検体名、診断キー、診断名、TNMキー、TNM分類、Stageキー、Stage、指摘キー、誤字やStageの間違いの指摘、転移キー、転移値(5つ)、バイオマーカーキー(6つ)、バイオマーカー値(6つ)

事象	抽出臨床項目	項目数 (採点結果)	備考
2019年12月初旬に胸痛、背部痛を自覚、近医を受信し左肺門部腫瘍を指摘される	日時、症状	2	色つきが全て抽出できているもの
1月7日に当院を紹介受診。同日に胸腔穿刺を施行し、得られた胸水から肺腺癌と診断 T4N1M1c StageIVC, BRA, PUL, LYM, PLE, OSS, EGFR(L858R+), KRAS-, BRAF(V600E)-, ROS1-, PDL-1<0%, ALK-	診療日、検体キー、検体名、診断キー、診断名、TNMキー、TNM分類、Stageキー、Stage、指摘キー、誤字やStageの間違いの指摘、転移キー、転移値(5つ)、バイオマーカーキー(6つ)、バイオマーカー値(6つ)	30	
1/31~:C1-7頸椎転移に対し緩和的放射線治療(30Gy/10fr) 施行	開始日、治療キー、治療内容、部位キー、部位名、線量キー、線量値	7	
2/6~: 1stline Osimertinib 80mg/dayを開始(28日分)。	開始日、ラインキー、治療ライン、治療キー、治療タイプ、薬剤キー、薬剤名、投与キー、投与量、期間キー、投与期間	9	
11/4: CTにて左肺門部腫瘍、左肺尖結節、肝転移、骨転移増大を認めPDと判断。	診療日、判定手段、検査内容、判定内容	4	
11/19~2/10: 2nd line CBDCA/PTX/Bev/Atezoを4course施行	開始日、終了日、ラインキー、治療ライン、治療キー、治療タイプ、薬剤キー、薬剤名、投与キー、投与量	8	
2021/3/2: 効果判定にて原疾患の増悪を認めPDと判断。	診療日、判定手段、検査内容、判定内容	4	
3/16~4/13: 3rd line DOC+RAM 2course施行	開始日、終了日、ラインキー、治療ライン、治療キー、治療タイプ、薬剤キー、薬剤名、投与キー、投与量	8	
5/18:効果判定にて両肺の小結節は増加・増大と両側胸水の増加がありPDと判断	診療日、判定手段、検査内容、判定内容	4	
5/20: 胸水コントロール目的に入院。	診療日、治療キー、治療内容	3	
5/21: 左瘻性に対して胸水左胸腔ドレーン挿入	診療日、治療キー、治療内容	3	
5/24: 左胸膜癒着術(ユニタルク4g)を施行	診療日、治療キー、治療内容	3	
5/27~: 4th line EGFR-TKI rechallenge(Afatinib 20mg/day)開始	開始日、ラインキー、治療ライン、治療キー、治療タイプ、薬剤キー、薬剤名、投与キー、投与量	7	
抽出項目数の合計		100	92

- 臨床現場では700億パラメータをもつLLMを使って生成するのは計算資源の問題から難しい
- 量子化モデルが現場でも利用可能かを調査
- 各癌の症例を3回評価して平均をとった
- 対象

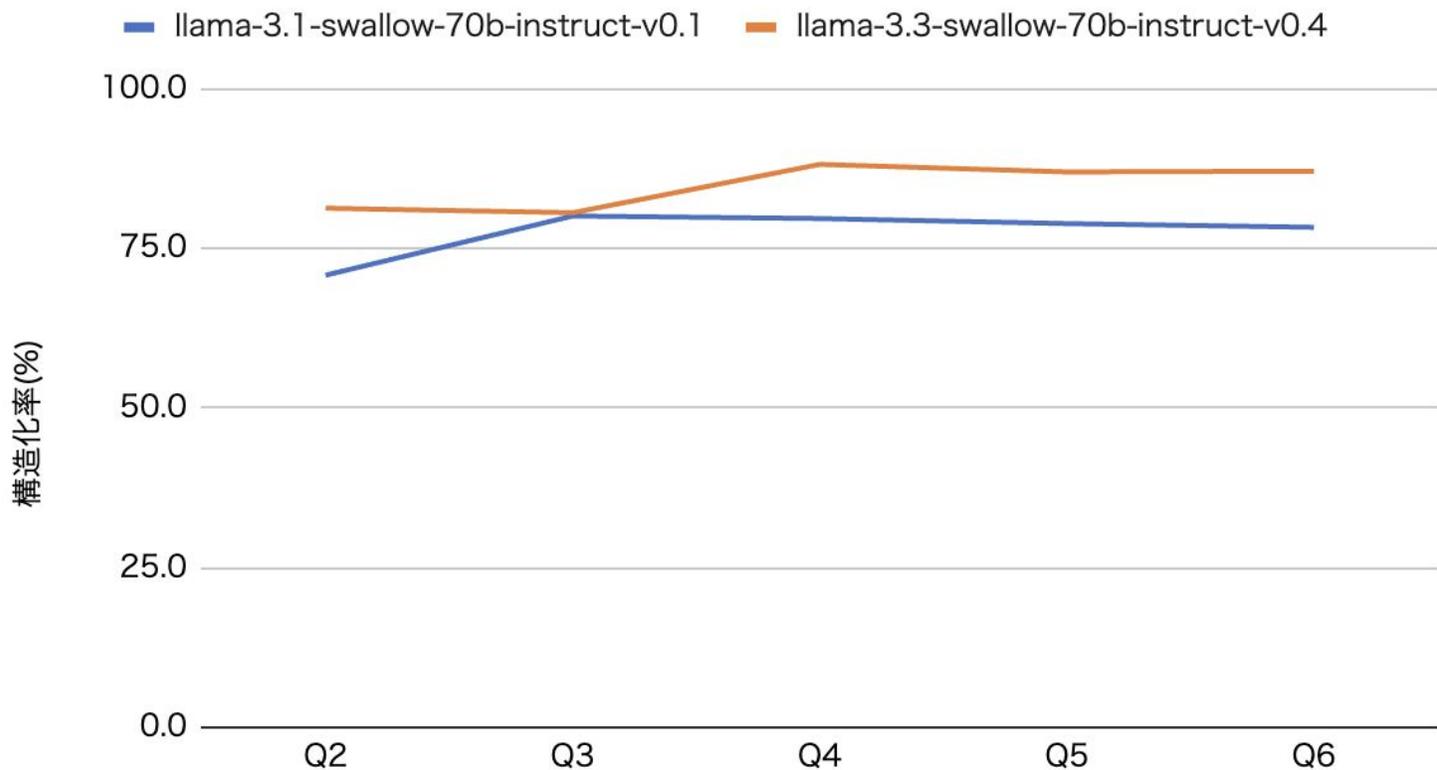
- 胃がん
- 前立腺がん
- 大腸がん
- 乳がん
- 肺がん

## 量子化影響調査



Llama3.1-Swallow-70B-Instruct, Llama3.3-Swallow-70B-Instruct  
構造化精度比較グラフ

## 量子化と構造化率



Llama3.1-Swallow-70B-Instruct,  
Llama3.3-Swallow-70B-Instruct構造化精度比較表

	llama-3.1-swallow-70b-instruct-v0.1	llama-3.3-swallow-70b-instruct-v0.4
Q2	70.8	81.3
Q3	80.1	80.6
Q4	79.7	88.2
Q5	78.9	87.0
Q6	78.3	87.1

1

医療用LLM

2

構造化用LLM開発

3

まとめ

