

# ABCI 3.0開発加速利用（2025年度）成果概要（公開用）

課題名：AIの安全性確保に関する研究開発・検証等の推進事業／日本語版医療特型LLM化の社会実装に向けた安全性検証・実証

実施時期：令和7年度  
 所属機関名：さくらインターネット株式会社  
 代表者氏名：小西史一

成果概要：医療機関がLLMを活用する際、クラウドLLMへの患者情報送信によるデータ主権リスクが構造的課題となっている。本事業では、患者データを院外に出さずに利用できるフロンティアモデル級の国産医療LLMの構築を目指し、データ調達、モデル開発、安全性検証、ユースケース検証の4項目を実施した。独自収集のデータを用いて高性能かつ安全な日本語版医療LLMおよび周辺システムを開発後、実際の医療機関データを用いて、6情報の変換・3文書の生成・レジストリの生成のユースケース検証を行った。

成果のポイント：ABCIでは、東京科学大学のチームが、Swallowコーパスを活用した汎用追加学習モデル（Qwen3-Swallow系、GPT-OSS-Swallow-120B）および医療特化モデル（medical-qwen系、Medical-GPT-OSS-Swallow-120B）を開発した。

モデル	開発組織	サイズ	医学系タスク				安全性		汎用性能	
			医師国試	専門医	専門医 (RAG)	ガイドライン	対話型診断	DNA	AC	NB4
<b>フロンティアモデル</b>										
Claude Opus 4.5	Anthropic	—	<b>99.1%</b>	90.6%	91.4%	<b>69.8%</b>	44.0%	<b>97.3%</b>	<b>90.8%</b>	<b>96.1%</b>
GPT-5.2	OpenAI	—	98.8%	91.2%	—	67.8%	<b>60.9%</b>	95.3%	<b>90.8%</b>	95.4%
Gemini 3 Pro Preview	Google	—	98.0%	<b>92.5%</b>	93.3%	68.0%	46.9%	95.3%	89.6%	<b>96.1%</b>
<b>追加学習モデル</b>										
Weblab-MedLLM-GLM-4.7	東大 (追加学習)	355B	<b>95.9%</b> (-1.5)	<b>83.5%</b> (+1.6)	<b>90.8%</b>	<b>63.7%</b> (+10.8)	45.1% (-0.5)	96.6% (-3.0)	<b>89.0%</b> (-9.5)	90.2% (+1.2)
Weblab-MedLLM-Qwen3-235B-Thinking	東大 (追加学習)	235B	95.7% (±0.0)	82.7% (+1.7)	89.1%	60.6% (+4.1)	51.5% (-1.6)	<b>97.0%</b> (-0.2)	88.4% (-3.0)	89.1% (-6.4)
Weblab-MedLLM-gpt-oss-120b	東大 (追加学習)	120B	93.6% (+3.8)	78.0% (+2.1)	88.6%	45.6% (-7.7) ※3	<b>54.0%</b> (+4.6)	96.4% (+1.3)	86.9% (+1.4)	<b>94.0%</b> (+16.2)
Medical-GPT-OSS-Swallow-120B	東京科学大 (追加学習)	120B	87.9%	66.5%	—	—	—	92.6%	78.6%	—
Medical-Qwen3-Swallow-32B	東京科学大 (追加学習)	32B	85.4%	58.9%	—	—	—	94.1%	76.2%	—

## 九大データを用いたレジストリ生成精度の比較

モデル	系統	種別	全体正解率
Weblab-MedLLM-GLM-4.7	<u>weblab</u>	追加学習	<b>45.8%</b>
GLM-4.7	—	ベース	44.3%
Qwen3-235B-A22B-Instruct-2507	—	ベース	44.0%
Qwen3-32B	<u>weblab</u>	ベース	42.6%
Weblab-MedLLM-Qwen3-235B-Instruct	<u>weblab</u>	追加学習	42.5%
Qwen3-235B-A22B-Thinking-2507	—	ベース	41.9%
Medical-Qwen3-Swallow-32B	<u>weblab</u>	追加学習	41.8%
Weblab-MedLLM-Qwen3-235B-Thinking	<u>weblab</u>	追加学習	40.5%

成果についてより詳細な情報を提供しているWebページ、発表論文などの情報：