

テーマ4：医療データ・医療LLM/LMMの利活用を促進する 医療データ基盤

◎中野 美由紀, ◎田村 孝之, 合田 憲人, 高倉 弘喜, 林 正和, 伏見 信也, 佐藤 健
情報・システム研究機構

研究の目的

我が国の医療LLM/LMMの研究開発を促進

- 医療データを収集・加工・蓄積し、研究者が利用可能とする**医療データ基盤**を構築
- 医療データ基盤上のデータを安全かつ適正に利活用するための**仕組み**を考案

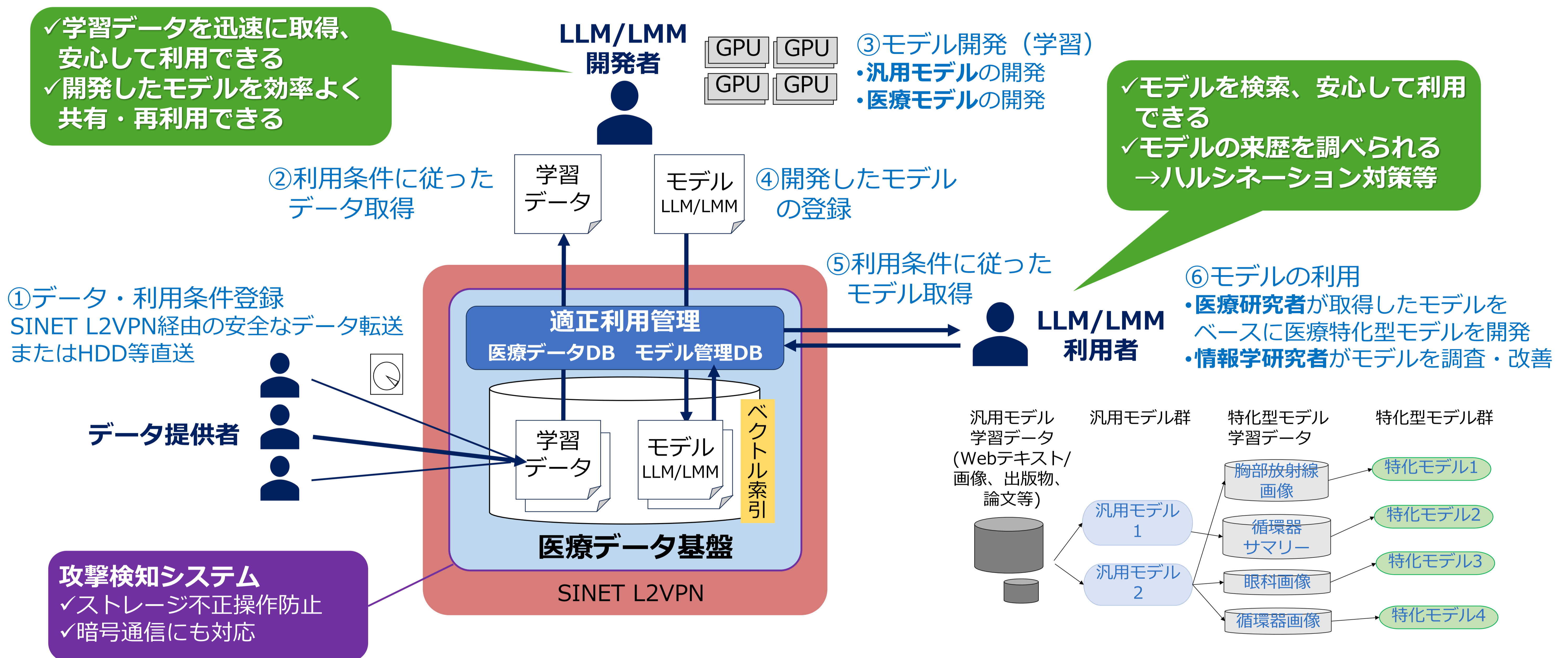
研究開発を促進するために解決すべき課題

- 大規模医療データ基盤
 - 大規模医療データの収集・管理が困難
 - 多様なLLM/LMMモデルの管理が必須
ハルシネーション等の技術課題解決への支援が必須
- データ収集・持続的な利活用
 - 医療LLM/LMM開発のための体系的なデータ収集が困難
 - 医療データの**適正利用**のための仕組みがない
 - 法・倫理的課題**の解決が必須

管理対象データ一覧（テーマ1, 2, 4全体）

データカテゴリ（分類）	件数・量	
クローラデータ		
テキスト	2.12兆トークン	
画像・テキストペア	18億ペア	
高品質医療テキスト		
医学書院教科書 医中誌Web	17.6億文字	
医療データ	連携先	
SIPテーマA-1参加13施設(CLIDAS)からのデータ等（DPCLセプト、心電図・心エコー数値、胸部X線画像）	九州大学 東京大学	3.8万件
日本医用画像データベースのCT・MR画像	順天堂大学 NII	5.6億枚
特化型LLM/LMM学習用データ	自治医科大学 東京大学	119万件

医療データ基盤の利用イメージ



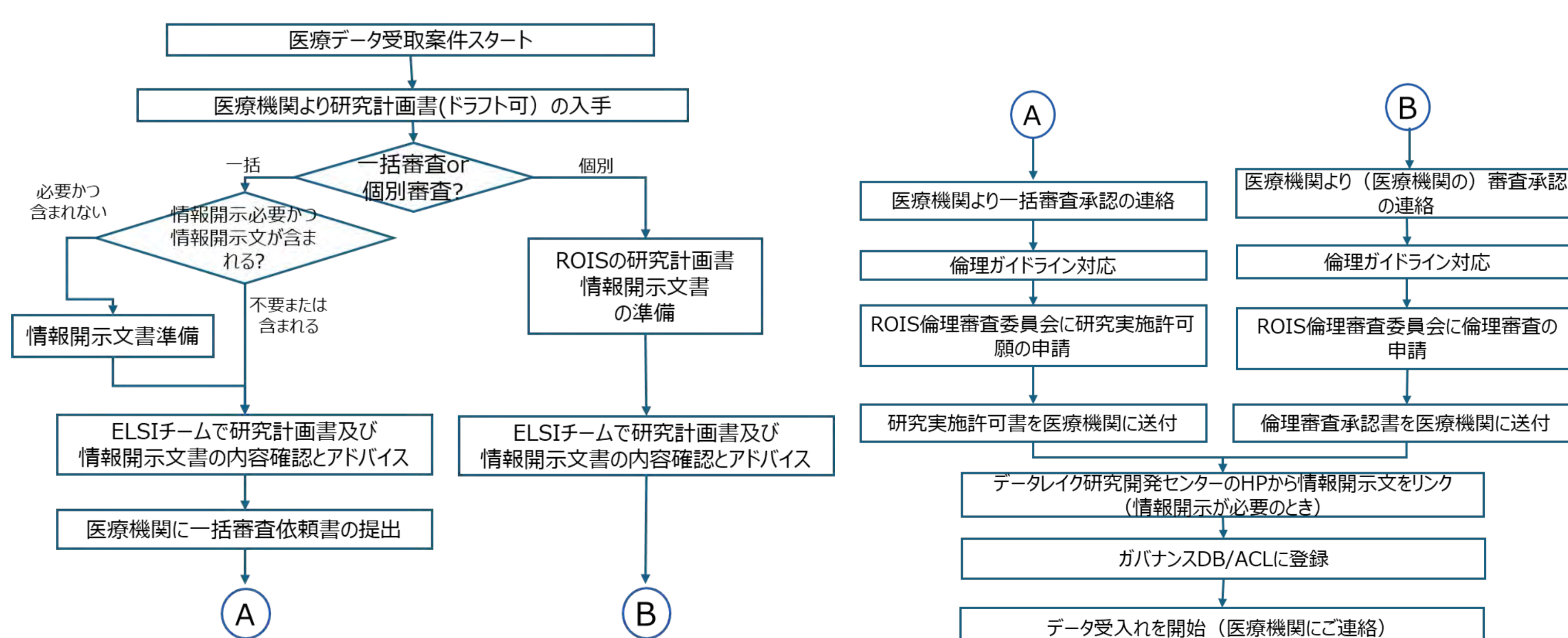
ELSIグループ設置による法的・倫理的課題の検討

検討結果や提言を報告書（700ページ）として**公開予定**

<https://ds.rois.ac.jp/center8/>

医療データ受入プロセス

倫理ガイドラインに沿ったデータ受入のためのプロセスを
ELSIグループの支援を得て策定し実践



検討事項	成果
生成AI開発フェーズ での法的・倫理的課題の検討	<ul style="list-style-type: none"> 法・倫理指針に則ったデータ収集・提供の枠組みを整理・実運用 著作権法：非享受目的での利用・著作権者との個別契約 個人情報保護法：学術研究例外の適用 人を対象とする生命科学・医学系研究に関する倫理指針
研究途上 において生じる課題の検討とフィードバック	<ul style="list-style-type: none"> テーマ1・2の研究者への法・倫理に関する支援：チュートリアル、アドバイス 医療テキスト出版社・団体との契約支援 データ受け入れ手順の作成・運用支援
生成AI利活用フェーズ での法的・倫理的課題の検討	<ul style="list-style-type: none"> 著作権法・個人情報保護法：法的視点からの侵害を防ぐ技術の必要性を開発へフィードバック 仮名加工情報、社会実装を前提とした法的枠組み整理（次世代医療基盤法、共同利用）
諸外国の制度の整備状況 についての情報収集と予想される国内法規制への対応検討	<ul style="list-style-type: none"> 海外調査：各国の医療生成AIに関する制度・動向調査（EU、米、デンマーク等） 政策提言の作成：医療データ活用、安全性確保、AIガバナンスの視点等から提言を整理し、報告書と共に公開