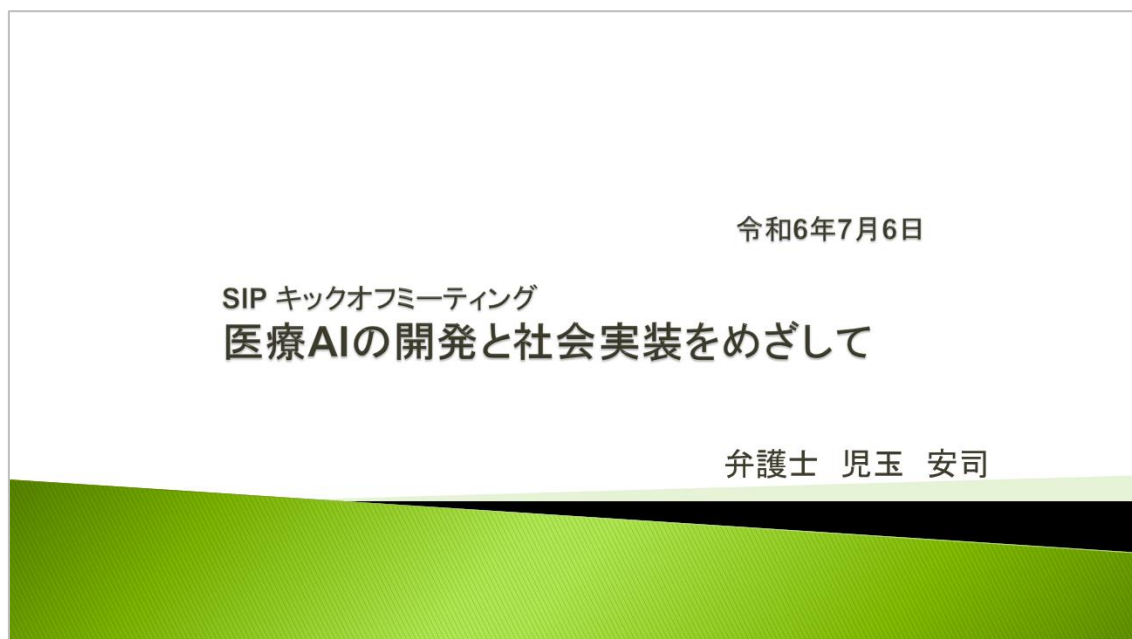


第1回 医療AIの開発と社会実装をめざして

(弁護士 児玉 安司【ゲスト】八木 隆一郎 (Brigham Women's Hospital 研究員)、原 聖吾 (株式会社 MICINM代表取締役)、合田 憲人 (ROIS・戦略企画本部))

はじめに



わが国では、医療データの利活用の重要性や方策が、実に長年議論されてきています。直近では、COVID-19 について、実際の臨床現場のデータが、諸外国では informed consent を取らない without consent で、直接、莫大に集められ、例えば、効果の調査であれ、副作用の調査であれ、急速に進められました。しかし、日本では様々な個人情報の利活用の壁があり、それを一つの要素として、わが国では COVID-19 に関連した主要な論文はゼロ、ワクチンはゼロ、治療薬はゼロが続いた最大の原因は、岩盤規制にあるというようなことが指摘されるようになっていきます。

しかし、その中で、次世代医療基盤法という個人情報保護法の特別法が、匿名加工医療情報というものから、直近の改正で、仮名加工医療情報へと展開し、使い勝手をよくしようとするなど、政策的な模索が続いています。

OUTLINE

1 アメリカにおける医療AI研究

2 このSIPにおける国内法の課題

3 SaMDからOpenAIへ

4 賢い患者(Smart Patient)という夢

(補1) GDPRとpublic health例外

(補2) ICMRA Horizon Scanning Assessment Report

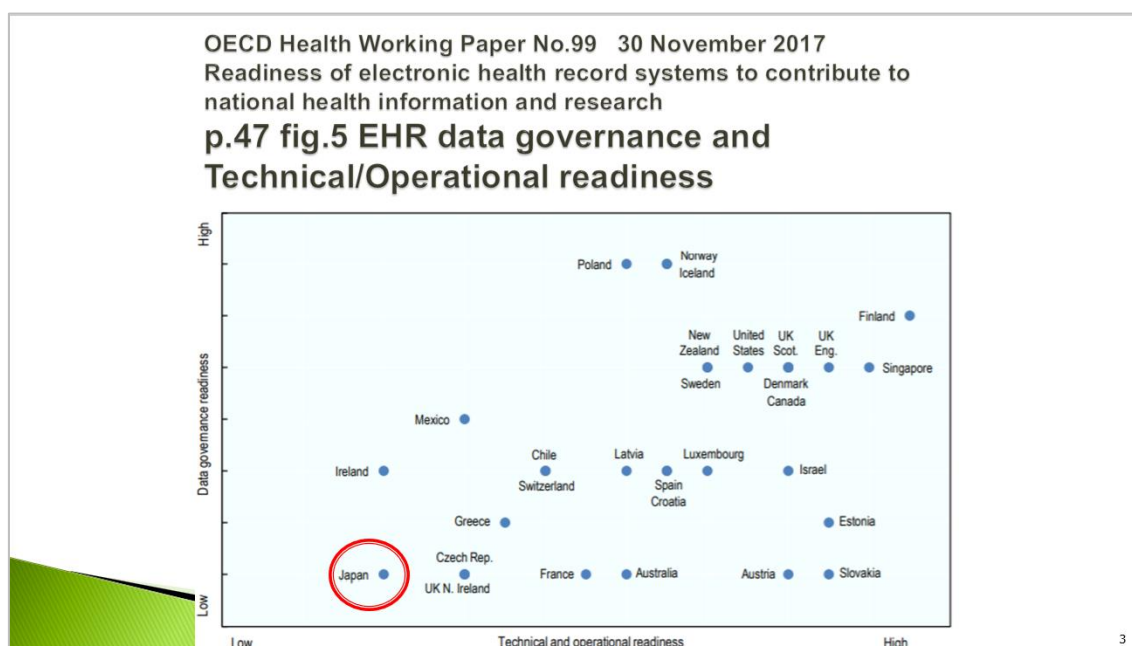
まず、アメリカにおける医療 AI 研究の具体例をお話します。

次に、この SIP における国内法の課題について、著作権法、個人情報保護法、製造物責任法、さらに医薬品医療機器等法、いわゆる薬機法に関わる医療機器としての承認の問題などについて説明します。

続いて、SaMD (Software as a Medical Device) です。AI 開発の最初に得られる果実がソフトウェアであるわけですが、このプロジェクトでは、ソフトウェアを作るところを超えて、医療の生成 AI をオープンにする、場合によっては OpenAI まで目指そうというようなビジョンがあります。

最後は、患者さんの情報と権利について、少し海外、特に中国などを含めてされている議論を紹介します。

GDPR やヨーロッパの ICMRA の「Horizon Scanning Assessment Report」などは、パワーポイントでは付けていますが、今日は時間の都合で割愛します。次回のミーティングで、日本の個人情報保護法制、GDPR の医療に関する条文について、今日はアメリカの HIPAA について詳しく説明します。



これは、2017 年に出された OECD の「Health Working Paper No.99」です。縦軸に個人情報の取得、収集、取り扱いという意味でのデータ・ガバナンスの準備状況（readiness）がどれだけ進化しているか。横軸に医療情報（EHR：Electric Health Record）の技術・運用面の準備状況（technical operational readiness）を取っています。OECD での順位を並べていくと、フィンランドやシンガポールは結構進んでいます。US も UK も、技術的にも制度的にも非常に進化していると評価されています。しかし、日本は、2017 年のレポートでは技術的にも制度整備でも OECD 最下位で、アイルランドやメキシコにも到底追いつけそうにないという、衝撃的なデータが発表されています。（なお、2020 年に、同じ Oderkirk 氏をリサーチャーとする報告書が OECD から出されており、結果がかなり変動していますので、注意を要します。ただ、医療 DX

を推進しなければならないという国をあげての認識の出発点となったという意味で、2017年のOECDリサーチペーパーは重要です。

「次世代医療基盤法検討WG 中間とりまとめ」（令和4年6月3日）のポイント	
1. 医療研究の現場ニーズに的確に応える匿名化のあり方の検討 ＜匿名加工医療情報では対応できない研究現場のニーズ＞ ① 希少な症例についてのデータ提供 ② 同一対象群に関する継続的・発展的なデータ提供 ③ 薬事目的利用の前提であるデータの真正性を確保するための元データに立ち返った検証 ○ 次世代法の認定事業者と利活用者におけるデータ・ガバナンスを強化することにより、提供先での匿名性は維持しつつ、有用性の高いデータを提供できるような匿名化のあり方を検討する。	2. 多様な医療情報との連結・収集 (1) NDBなど既存の公的データベースとの連結に向けた検討 ○ 診療報酬請求明細書（レセプト）を皆悉性高く把握できるNDBと連結解析できるよう検討する。 ※ NDBとの連結解析により、例えば、次世代法認定事業者がデータを保有している病院への受診（入院）前後に、他の診療所等でどのような受診をしたか把握でき、より精緻な研究開発が可能となる。 (2) 急性期病院以外の医療機関や自治体等のデータ収集の促進 ○ 医療機関や医療保険者等に対して、医療情報の提供について検討を促す方策を検討する。 ○ 質の高い疾患別レジストリを持つ学会や、健診情報などを持つ自治体などへの周知強化を検討する。 ○ その他、医療機関におけるオプトアウト通知の方法について、運用面の工夫により医療機関の負担軽減を図ることや、認定事業者によるデータカタログ開示の促進なども検討する。

個人情報保護法の枠を超えた医療情報に関する特別法である次世代医療基盤法という法律を作って、従来は匿名加工情報、つまり、元々のカルテとの関係を完全に切断してしまう匿名化を施した情報だったため、使い勝手が悪いという指摘がありました。

実際の活用例が少ないということで、「匿名」を「仮名」にする法改正がなされました。

実は、法政策を作る側の着想と、技術者側、特にデータサイエンス研究者の側の、プライバシーをどう守るかという、いわば暗号化技術のようなものの進化に、大きなずれがあります。このSIPの中では、患者のプライバシー保護をどう図っていくかということについて、実はひょっとしたら新しいモデルが提供されるかもしれません。匿名、仮名のような情報黒塗り戦略とは違ったかたちでのプライバシー保護の技術が、いま、実装できるかもしれないというのも、このSIPの大きなポイントの一つです。

1. アメリカにおける医療 AI 研究

1 アメリカにおける医療AI研究 US — HIPAA Health Insurance Portability and Accountability Act

アメリカの場合は、Health Insurance Portability and Accountability Act (HIPAA) という法律が、医療情報、とりわけリサーチについて明瞭なルールを提供しています。この HIPAA に基づいて、実際、どんなふうに医療情報が AI になっているのか。その実態について、八木隆一郎先生のお話を、4 分 30 秒ですので、聞いていただこうと思います。

▶【AI研究者の立場から】八木 隆一郎 医師 （4分30秒）

動画を放映いたします。

2021年11月18日 日本医師会生命倫理懇談会

医療分野におけるAIの現状

研究の立場から

八木隆一郎



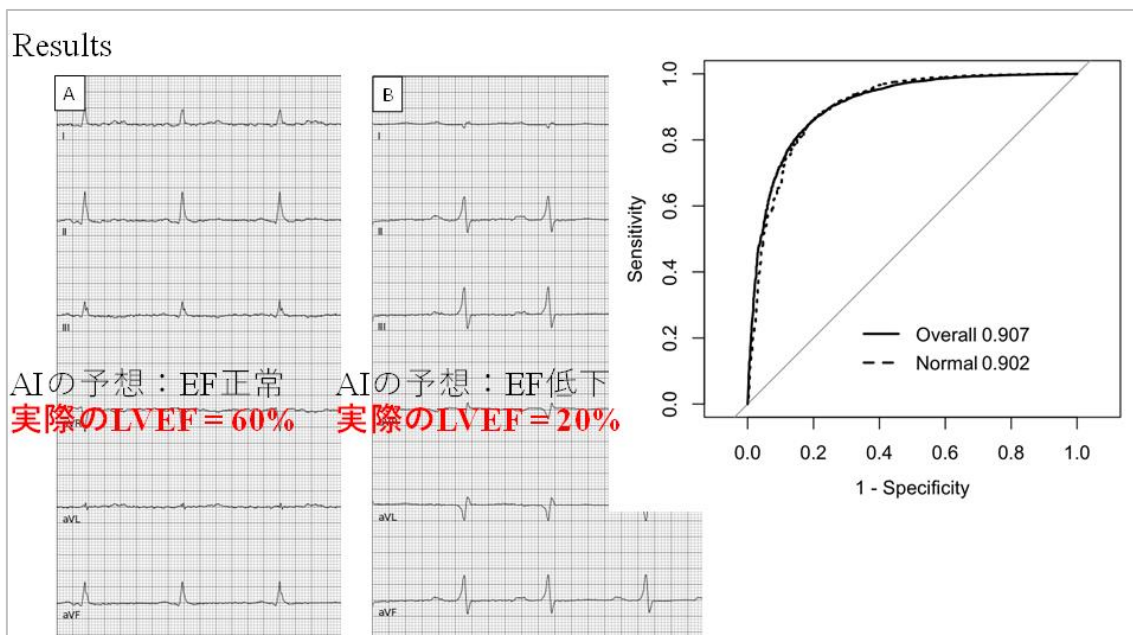
実際の研究

心電図から左心機能低下症例を検出する(約50,000症例)

AIは、心電図を見るだけでエコーの結果を予測し左室収縮能低下しているか判断できるか？



【八木隆一郎氏】 実際の研究ですが、心電図から左室の心機能低下をきたしている症例を検出するような AI の開発に取り組んでいます。心エコーとひもづいた心電図の画像データ約 5 万件を電子カルテから集めて、ディープラーニングで解析しました。



こちらは、A、B ともに似たような、よく見るような心電図ですが、AI に読ませると、A の心電図は EF 正常、B の心電図では EF が低下しているのではないかと予想できます。

【児玉】 コメントを入れますが、普通の心電図を二つ並べていて、EF というのは ejection fraction です。心電図は心臓の電氣的な動きを捉えていて、ポンプとしての機能は捉えきれていないと信じられていました。ejection fraction というのは心臓からどれだけの血液が駆出 (eject) されるか、その割合、駆出率を示すものですが、心電図の伝記的情報からポンプとしての動力学的な予測ができるのではないかという仮説の下に、研究を進められたわけです。

【八木隆一郎氏】 実際のところ、カルテを確認しますと、A の症例では EF60%と正常ですが、B の症例では EF20%と大きく低下していることが分かりました。

これは 1 例ですが、全体でも、AUC0.9 と高精度です。つまり AI は人間の医師では分からないような微小な心電図の変化や異常を捉えて、心電図を予測しているのではないかなんとも言えます。これは例えば、健診などで、心電図で明らかな異常は

ないが、心エコーをやったほうがいいのか、専門機関に紹介したほうがいいのかな、と悩むようなとき、こういった AI が実装されると診療の助けになると考えています。

研究に使ったデータは全て電子カルテから

病院への申請
個人情報に関する 1 時間ほどのビデオ講座



病院が研究者に対して匿名化されていない電子カルテデータの包括的な使用を許可
患者ID、年齢、性別、人種、受診日、診断名、
検査結果、画像データ、テキストデータ etc...

特徴

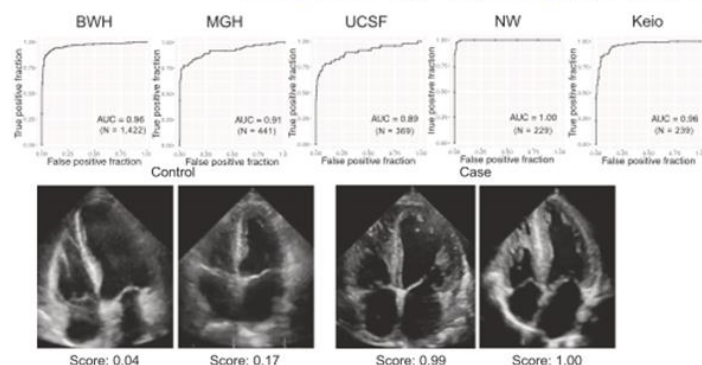
- ✓ 2014 年から・約 20 病院のデータ (ECG だけで約 3000 万人分)
- ✓ 個別の研究に対する患者の同意は不要 (オプト・アウト)
- ✓ 個別の研究に対する研究者から病院への申請も不要
- ✓ 自分の PC で自宅からアクセスも可能

大規模な研究が可能だった理由は、ハーバードには関連病院、20 病院のデータが 10 年分程ありまして、心電図だけで約 3000 万人分のデータが蓄積されていて、病院に申請を出せば、匿名化されていない電子カルテのデータそのものにアクセスやダウンロードが可能となっているからです。このシステムを使えば、許可を受けた研究者は研究に必要なデータを簡単にダウンロードでき、個別の研究に対する病院や倫理委員会の許可や、患者さんの個別の同意などは不要で、思いついたらすぐに研究ができて、リモートでも解析可能という特徴もあります。このような電子カルテデータを使った AI の研究は、症例を集めるのが難しい希少疾患の研究にも応用できます。

心エコー動画からAIを用いて心アミロイドーシスを診断できるか？

Artificial intelligence-enabled fully automated detection of cardiac amyloidosis using electrocardiograms and echocardiograms

Shinichi Goto^{1,2,3}, Keitaro Mahara⁴, Lauren Beussink-Nelson⁵, Hidehiko Ikura³, Yoshinori Katsumata³, Jin Endo³, Hanna K. Gaggin^{2,6}, Sanjiv J. Shah⁵, Yuji Itabashi³, Calum A. MacRae^{1,2} & Rahul C. Deo^{1,2,7,8}



約3000症例を解析
(うち慶応病院から52症例)

我々の研究室の研究ですが、診断に難渋することが多い疾患である心アミロイドーシスを、AI を用いて単純にエコー動画から診断できないかという後ろ向き研究です。AI はエコー動画を見るだけで高い精度で正しく診断できたと報告されています。この研究では、心アミロイドーシス 3000 症例の情報を電子カルテから集めてきています。

電子カルテデータとAIの組み合わせ

✓今まではできなかった、臨床に直結する研究が可能に

データが使えるれば日本でもできる研究

✓むしろデータの質自体は日本の方が高い？

臨床上の複雑なタスクは、いくら AI といえども大量のデータのインプットを必要としますが、既にある電子カルテデータと AI を組み合わせることによって、患者さんや先生方の診療にダイレクトに役立てられるような研究が可能となってきています。

先の研究は、特殊な設備や機械を使っているわけではなくて、データさえあれば日本でも行える研究です。日本は診療の質が高い、検査の質が高いと感じます。例えばエコー動画一つとってみても、こちらのエコー動画はちゃんと切れていなかったり、見たいビューがなかったりということがありますが、日本では基本的にそういうことはありません。単純に、データベースの質は診療や検査の質がそのまま反映されるため、病院の電子カルテ群のデータが包括的に使用可能となれば、高いクオリティとなって非常に魅力的な研究環境となると考えています。

【以下、児玉】アメリカで現に行われていることには、患者の生の電子カルテデータに実際にログインして、自宅からでも研究を続けられるという利便性があります。日本の研究者から見ると、日本の電子カルテというのは、USB さえ入れることができないし、ネット環境から遮断するのが当然とされます。昨今、ランサムウェア等の問題が出てきて、セキュリティ技術の検討が進んでいるところです。患者さんの名前が付いたものを、直接研究の対象にしていくことはいわばタブーに近く、匿名化したり仮名化したりして、何か黒塗りしなければ研究ができないというのが、基本的に日本のこれまでの研究の立場なわけです。

かつてできていたことを、妨げない

▶ かつて

医学研究者は、何か思いついたら、まず病院内の患者さんのカルテを研究目的で読んでみる

▶ 現在

医学研究者は、何か思いついたら、まずビッグデータにアクセスして研究目的でAIに読ませてみる

情報インフラの形成
セキュリティの確保
手続・ルールの整備

アメリカでは、何千万人分ものデータが使用できるのかということ、カルテが電子化される前、紙の時代にできていたことを、電子化されたビッグデータの時代にも妨げないというプリンシプルが、アメリカの政策の根幹にあるのです。医学研究者は、何かを思いついたら、まず病院内の患者さんのカルテを研究目的で読んでいました。治療目的で来た患者さんのカルテを、似たような患者がどこかにいたな、似たような心電図、エコーはないか、というようにどんどん調べてみます。そのように手作業でカルテを探して研究のアイデアを探していた時代から、ビッグデータにアクセスして、まず AI に読ませてみるという時代になってきているわけです。

そのためには、医療情報のデータ基盤、情報インフラを形成しなくてはなりませんし、技術的な核心部分のセキュリティ確保、それからプライバシー保護、そのための手続きやルールをきちんと整備していかなければなりません。同じ患者さんに何度もアクセスするということは、その患者のことを研究していて、逆に、ロボットがその患者の心電図を 1 回だけ読んでいくのであれば、それは学術研究に違いない、アクセスのパターンで仕分ければ、プライバシー侵害をかなりの率で防御できるなどのアイデアがあるようです。

BEYOND THE HIPAA PRIVACY RULE (2009)

Enhancing Privacy, **Improving Health Through Research**

Institute of Medicine of the National Academies

▶ アメリカ学術会議 IOMレポート

Congress recognized the important role that health records play in conducting health research and wanted to ensure that privacy protections would not impede researchers' continued access to such data.

連邦議会は、健康に関する研究を行う上での健康情報の果たす重要な役割を認識し、プライバシー保護が**研究者のデータへの持続的アクセスを妨げてはならない**ことの保証を求めた。

この背景となる政策として、アメリカ学術会議が 2009 年に出した IOM レポート「Beyond the HIPAA Privacy Rule」というものがあります。「連邦議会は、健康に関する研究を行う上での健康情報の果たす重要な役割を認識し、プライバシー保護が研究者のデータへの持続的アクセスを妨げてはならないことの保証を求めた」。

研究者が、電子カルテというかたちで利用の汎用性が高まったビッグデータに持続的アクセスをすることで、医学研究が達成されるということです。

U.S. Congress, 1996

two House Reports on HIPAA (上院・下院同文)

The conferees recognize that certain use of individually identifiable information are appropriate, and **do not compromise the privacy of an individual.**

個人を特定できる情報のある種の使用は適切であり、**個人のプライバシーの侵害にはならない。**

ここに引用した連邦議会決議は、1996 年の上院・下院の決議ですが、「個人を特定できる情報のある種の使用は適切であり、個人のプライバシーの侵害にはならない」とあります。

U.S. Congress, 1996

two House Reports on HIPAA （上院・下院同文）

Example of such use include...

the transfer of information from health plan to an organization for the sole purpose of conducting health care-related research. As health plan and providers continue to focus on outcomes research and innovation, it is important that the exchange and aggregated use of health care data be allowed.

プライバシーの侵害にならない例

医療に関連する研究の実施のみを目的とする健康保険から研究機関への情報移転。健康保険と医療提供者はアウトカム・リサーチとイノベーションを追求し続けているのだから、医療データの情報交換と重層的利用が許されるべきである。

プライバシーの侵害にならない例について上院・下院の同文決議は「医療に関連する研究の実施のみを目的とする、健康保険から研究機関への情報移転、健康保険と医療提供者はアウトカム・リサーチとイノベーションを追求し続けているのだから、医療データの情報交換と重層的利用が許されるべきである」と述べています。

例えば、健康保険を使って重篤な疾患を治療したら、その人は COVID-19 のワクチンを打ったあとだったというような情報を、片端から集めてきてリサーチをすると、ワクチンの副作用情報が集まります。あるいは、ワクチンだけではなくて、さまざまな抗がん剤をはじめとした副作用の強い薬の情報も、このようにすればたちどころに集まってきます。

HIPAA

Research use and disclosures without individual authorization

個人の同意のない「研究」利用

The HIPAA Privacy Rule establishes the conditions under which protected health information may be used or disclosed by covered entities for research purposes. **Research** is defined in the Privacy Rule as, “**a systematic investigation, including research development, testing, and evaluation, designed to develop or contribute to generalizable knowledge.**” See 45 CFR 164.501.

Generalizable knowledge (一般化できる知識)をめざすものを Research(研究)として別扱いする

そういった人間の健康と医療の革新を支えていくために、この HIPAA という法律が作られましたが、この法律の中では、Research use and disclosures without individual authorization (個人の同意のない研究利用) というものを認めています。Generalizable knowledge (一般化できる知識) を目指すものを、research (研究) として別扱いします。

日本の場合は、学術研究機関が学術研究目的でやっていることを学術研究とするという、個人情報保護法の同意原則の例外規定が漸くできたわけですが、アメリカの場合はもう少し広く、学術という言葉にこだわらず、Generalizable knowledge (一般化できる知識) を目指すものを research として、倫理委員会の authorization によって、個人の同意がなくても研究できるということが、八木先生の研究を支える制度基盤になっています。

2. この SIP における国内法の課題

2 このSIPにおける国内法の課題

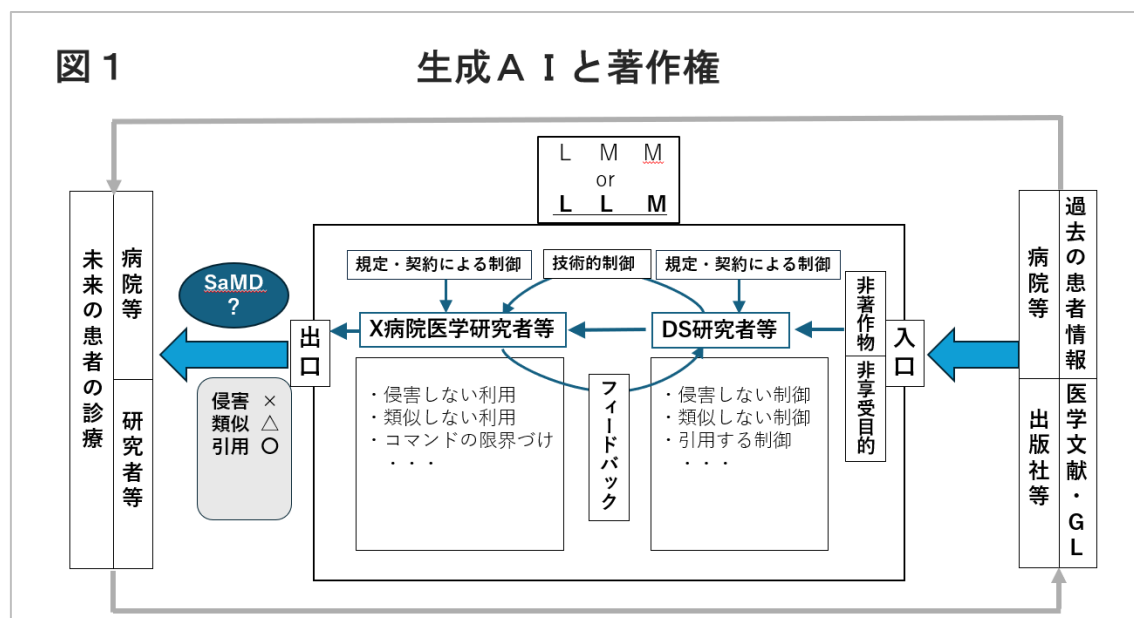
課題

- ▶ 医療生成AIと著作権
- ▶ 医療生成AIと個人情報保護
- ▶ 医療生成AIと製造物責任(薬機法の承認)

➡別添資料(20240706暫定版)参照

キックオフの段階で、まず出発点として、医療生成 AI と著作権の問題があります。今回、患者のカルテ情報も仮名化したかたちで AI に読ませますし、医学文献も AI に読ませようとしているため、著作権の問題が出てきます。それから、再三繰り返されている医療生成 AI と個人情報保護の問題。それから、出口での医療生成 AI と製造物

責任をどのように考えたらいいいのか。あるいは、医薬品医療機器総合法（薬機法）の医療機器としての承認を取るにはどうしたらいいのかといった課題が出ています。

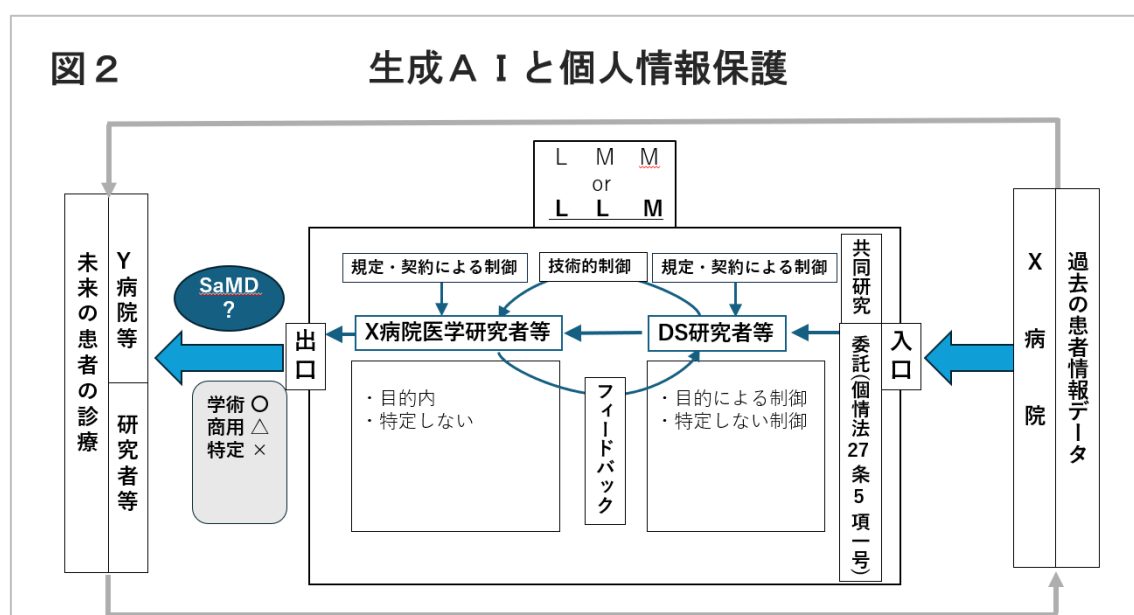


この図は、過去の患者情報が病院等から、医学文献・ガイドライン等が出版社等から、研究チーム・医療チームとデータサイエンスの研究者が、互いにフィードバックを繰り返しながら研究する学術研究の内側にいます。その入りの規制は、著作権法 30 条の 4 などでは緩めていますし、過去の患者情報等は著作物ではなくて、むしろ非著作物として取り扱うという考えに立つことができます。

また、軽微利用などの著作権法の例外規定等で、LMM (Large Modality Model) や LLM (Large Language Model) に文字情報や画像情報を読み込ませていく入り口のところは通過できます。研究者集団内での規定や契約による規律、それから濫用ができないようにする技術的制御などを工夫して、研究を進行させるところまでは、現状でも、現在の著作権法の枠組みの中で可能です。

問題は出口です。これを SaMD (Software as a Medical Device : 医療機器としてのソフトウェア) として出す段階では、例えば、心電図の機械にアルゴリズムを組み込んで、心不全を予測できる機械にするソフトウェアなどは、医学文献の中に出てくる表

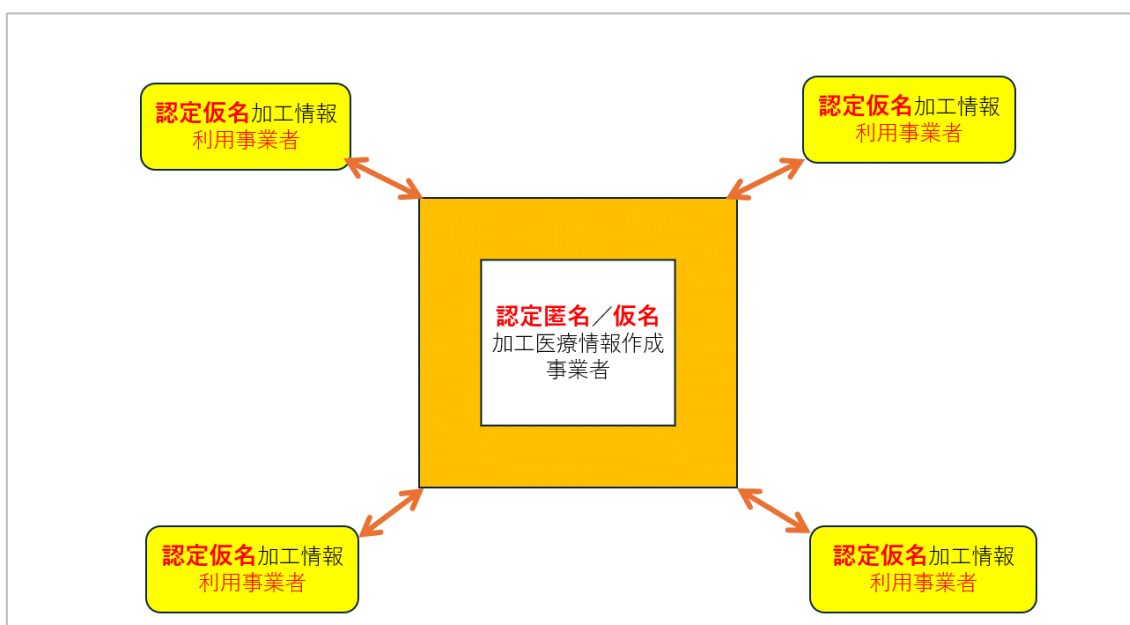
現がそのまま出てくることはないため、SaMD として出口から出ていく分には、それほど困難は起こりません。しかし、この LMM、LLM 自体をオープン AI として外に出すということになると、個人情報保護委員会からの注意喚起などに抵触する可能性があります。著作権については、今年の 3 月 15 日に出た文化庁の報告書が大変分かりやすく、生成 AI を作っていくに当たって著作権が阻害要因になることは、比較的少ないのではないかと考えています。



個人情報保護に関しては、例えば、X 病院のデータを入口から入れるために、情報回帰を委託するのであれば個人情報保護法 27 条 5 項 1 号の委託の条文を使う。データサイエンスの研究者と X 病院の医学研究者が、データを共同利用するのであれば、個人情報保護法 27 条 5 項 3 号の共同利用の条文を使い、定められた手続を履践する。共同の学術研究をするのであれば 27 条の 1 項 5 号、6 号、7 号の学術研究例外の規定を使う。このような整理をして、スタートしています。学術研究の共同研究の枠の中にあり、また委託の枠の中に入っているうちは、LMM、LLM とともに順調に成長していくだろうと予想されています。

しかし、その結果、出口からソフトウェアで出ていくとなると、ここから先は、そのソフトウェアの企業化の問題といった別の問題が新たに生じるわけです。

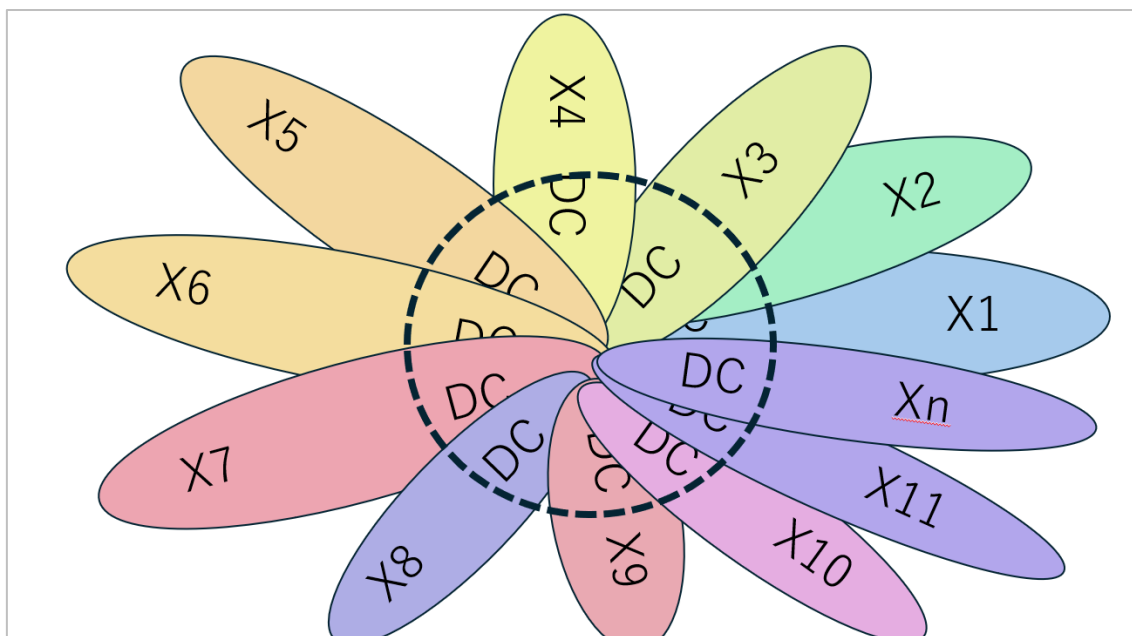
さらに、ここで得られた LMM、LLM を、個人情報の観点からどこまでオープンにできるのか。その場合、クエリーをどのように制御したらいいのか。また、著作権侵害にならないようにどのように制御していったらいいのか。などの課題が残っています。



次に、今回のこの SIP での個人情報の利活用は、委託や共同利用、共同の学術研究という枠の中で、適法ですが、旧来の医療情報全体については、内閣総理大臣、厚生労働大臣、文部科学大臣、経済産業大臣の 4 大臣から認定を受けて初めて認定匿名／仮名加工医療情報作成事業者になることができます。また、そうした事業者と情報をやり取りして利用する医療機関や研究機関の側も、認定仮名加工情報利用事業者としての認定を受ける必要があります。長期的なスキームとしては検討の余地がありますが、単年度の短期間でこの SIP を進行させていくには難しいと考えられました。

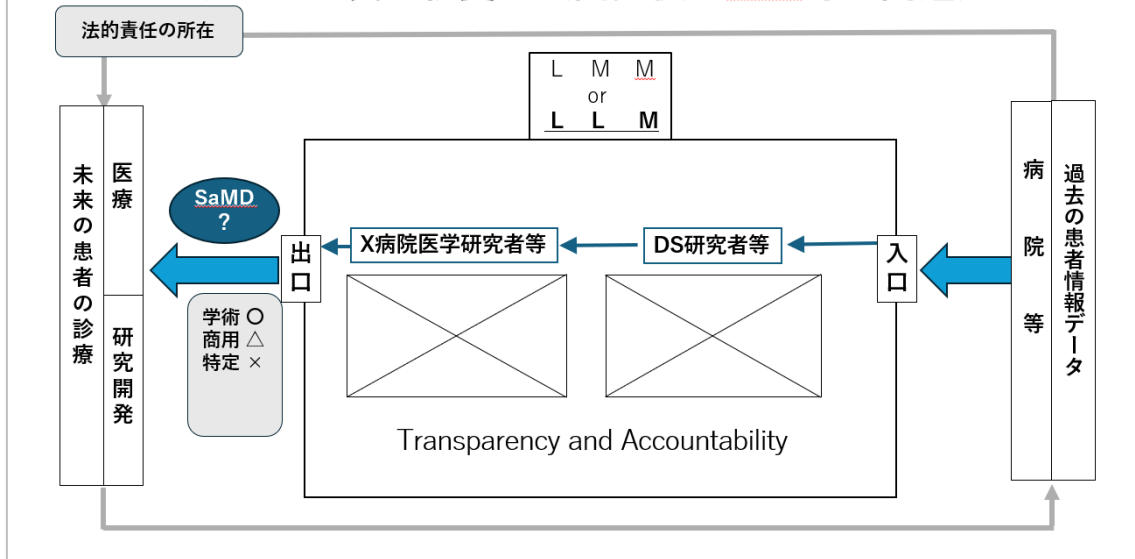
早急にデータセンターを作るのが難しければ、例えば、X1 病院で倫理委員会を通して既に行われている AI 研究や、医療情報を用いたデータ研究を、X 1 病院とデータ研

究機関が連携して、委託・共同利用・共同の学術研究などとして実施していくことは、倫理的にも法的にも問題ありません。



もし、いくつかの大学が、このように、それぞれ倫理委員会を通した臨床研究のプログラムを増やしていくと、真ん中に情報が集まっていくデータセンターらしき存在が出てきます。これを横に連結して本当のデータセンターにして、3000 万人分の心電図データというようなかたちに仕上げていくことがどこまでできるのか。それにはどんな制度構築が必要なのか。現行法の下で適法であり続けなくてはならないと同時に、政省令や国プロジェクトとして、個人情報保護法全体の枠組みの中で政策的によく考えていただかなくてはなりません。

図3 生成AIと製造物責任（開発後のSaMD等の問題）



それから、同じように入り口から出口まで、LMM や LLM で AI ができてくるとなると、生成AIと製造物責任ということで、AIに関する様々な倫理的・法的な規制が、5月22日のEUのAI Actなども含めて提案されています。共通しているのは、Transparency and Accountabilityを要求はするのですが、例えばEUのAI Actも、医薬品・医療機器の開発については規制サンドボックスのような適用除外があるようです。どのようにしてTransparency and Accountabilityを技術的に保証していくのか。また、SaMDの出口、SaMD以外の出口、このAI自体を拡張していくとなるとどういふことが起こるかという問題もあるわけです。

▶【起業家の立場から】原 聖吾 医師（7分17秒）

▶ 動画を放映いたします。

2021年11月

医療分野におけるAIの現状と課題 起業家の立場から

機密・専有情報
株式会社 MICIN による権利の明示的な承認を得ることなく、この資料を使用することを固く禁じます。

 MICIN

【原聖吾氏】 「医療分野における AI の現状と課題 起業家の立場から」ということで、情報提供します。

目次

1 会社概要

2 医療分野におけるAIの現状と課題

- ① データの蓄積・連携における課題
- ② データを活用したAIプロダクト開発における課題
- ③ プロダクトの臨床現場等への浸透における課題

【原聖吾氏】 本日の内容として、最初に、簡単に私どもの会社の紹介と AI の現状と課題について話します。

VISION

すべての人が、納得して生きて、最期を迎えられる世界を。

わたしたちは、テクノロジーを使って医療を変えていきます。

病院に眠っているデータを活用できれば、
ふだんの生活と病気の関係がもっとわかるようになる。

医師の"技"を"見える化"して伝えていけば、いつでも、
どこでも、誰でも、質の高い医療を受けられるようになる。

医療を身近にすることで、病気になった時に
「こんなはずじゃなかった」と感じる人をなくしたい。

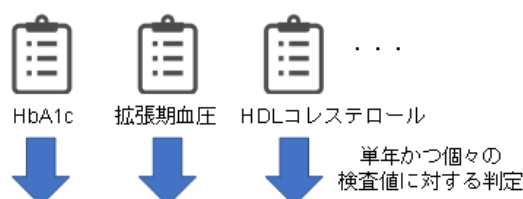
すべての人が納得して生きて、最期を迎えられる。
そんな世界を作っていきます。



MICIN という会社は、「すべての人が、納得して生きて、最期を迎えられる世界を」といったビジョンを掲げて、オンライン診療であり、医療データを活用した AI の取り組み、そういった事業に取り組んでいます。

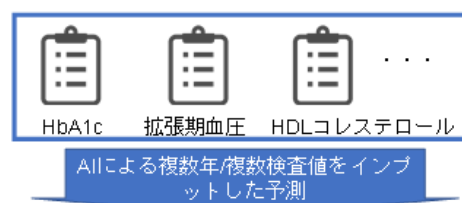
事例：健診結果からの疾病予測を開発し、健保組合や企業人事部長経由で個人に提供している

通常の健診結果：健康診断が“診断”のみで完結してしまっている



- 検査値単体での良し悪し判断
- 連続値としての処理がなされず、単年結果での判定結果からの所見のみを記載
- それが故に、特定保健指導の実施率も低水準

AIリスク予測：診断結果に基づく生活改善や早期受診に結びつく



- 属性や複数検査値、問診結果を組み合わせる個人別の健診値予測を実施
- 予測の個別化と精度向上により、生活改善や早期受診への動機づけが可能
- 個別の予測実施により、特定保健指導の実施率向上も期待

【原聖吾氏】具体的に医療 AI の取り組みの一つは、健診結果から疾病予測をするプログラムを開発して、患者自身、あるいは患者の手前の方の健診のデータから疾病を予測するというサービスを展開しています。

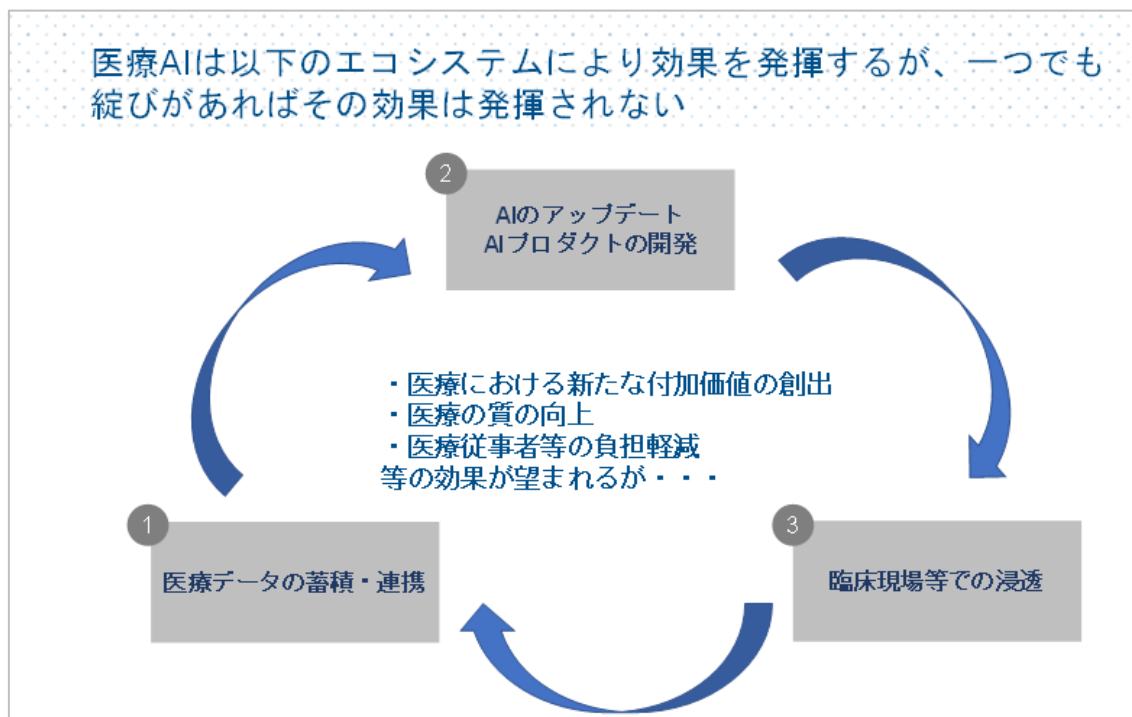
目次

1 会社概要

2 医療分野におけるAIの現状と課題

- ① データの蓄積・連携における課題
- ② データを活用したAIプロダクト開発における課題
- ③ プロダクトの臨床現場等への浸透における課題

【原聖吾氏】「医療分野における AI の現状と課題」と題し、特に次の三つの領域、データの蓄積・連携における課題、データを活用した AI プロダクト開発における課題、プロダクトの臨床現場等への浸透における課題、それぞれの点についてお話しします。



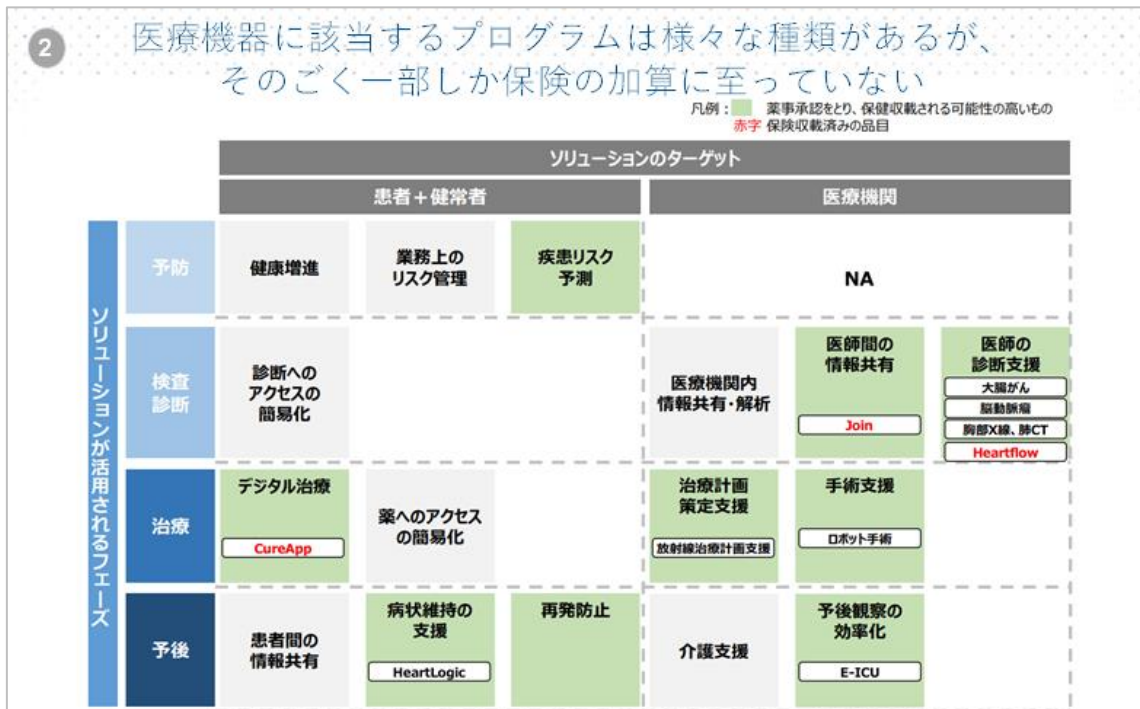
【原聖吾氏】まず AI というものは、図の①、②、③に書かれているもの、それぞれが組み合わさったかたちで、その成果を大きく発揮しています。まず①の蓄積・連携というポイント。そして、そのデータを基に AI プロダクトの開発、あるいは AI をアップデートするという②の部分。それが実際の臨床現場で使われるという③の部分。臨床現場で使われるとデータがより蓄積されていくということで、この三つがぐるぐると回って、さらに AI が進化していくところが、医療 AI の真骨頂であります。

このように考えてみたとき、それぞれの部分で課題があろうかと考えています。まずデータの蓄積・連携における課題について説明します。医療データ自体は、これまでも数としては非常に増えてきています。しかし、活用に当たってはこのようなステップごとに課題があると捉えています。

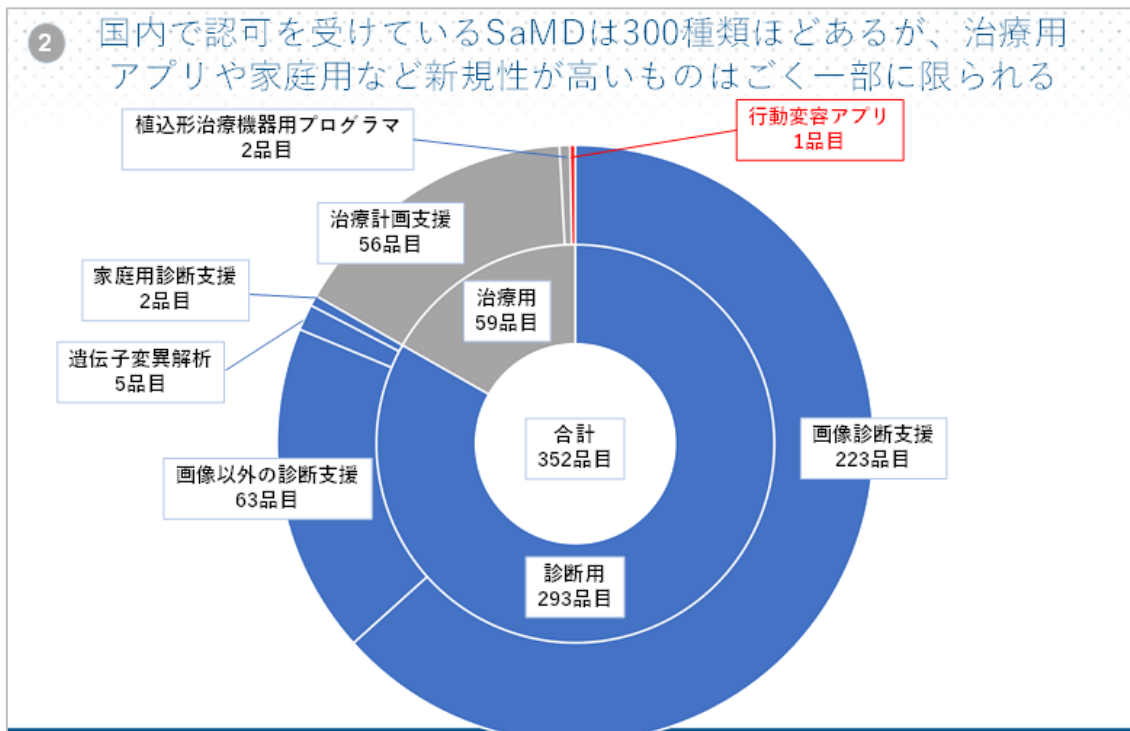
1 医療データ自体は増えてつつあるが、活用にあたっては複数のステップで課題がある	
事例・課題	
a データとして 収集できていない	・電子カルテ・オンライン診療等、医療におけるデータ化が進んでいない
b データはあるが、精度 にばらつきがある	・電子カルテやレセプトにおけるデータの信頼性が乏しい ・健康診断とレセプト等、IDが突合されていないデータがある
c データはあるが、商用 開発には使えない	・がん患者レジストリ・学会主導の医用画像データベース等は存在するが、商用開発に活用できない ・同意が取れていない、あるいは匿名化が難しい
d データ活用後の出口 戦略が描きづらい	・データを活用したビジネスの不確実性が高い

【原聖吾氏】まず、aとして、そもそも医療のデータが収集されていない、データ化されていないデータが様々あります。bは、データはあるが、その精度にばらつきがある。cは、データは実際にあり、一定の精度も保たれているかもしれないが、これを事業化していくという観点では使えないようなものになっている。最後、dとして、実際にこれが活用できるが、出口として、どのように活用するのかというところが見えづらい。こういった四つの観点で、データの蓄積というところに課題があります。

続いて二つ目のポイント、プロダクトを開発していくところの課題になります。医療AIを活用してどのようなサービス、プロダクトにしていくのか、様々な出口があります。その一つとして、医療機器のソフトウェア、いわゆる SaMD のような出口があります。



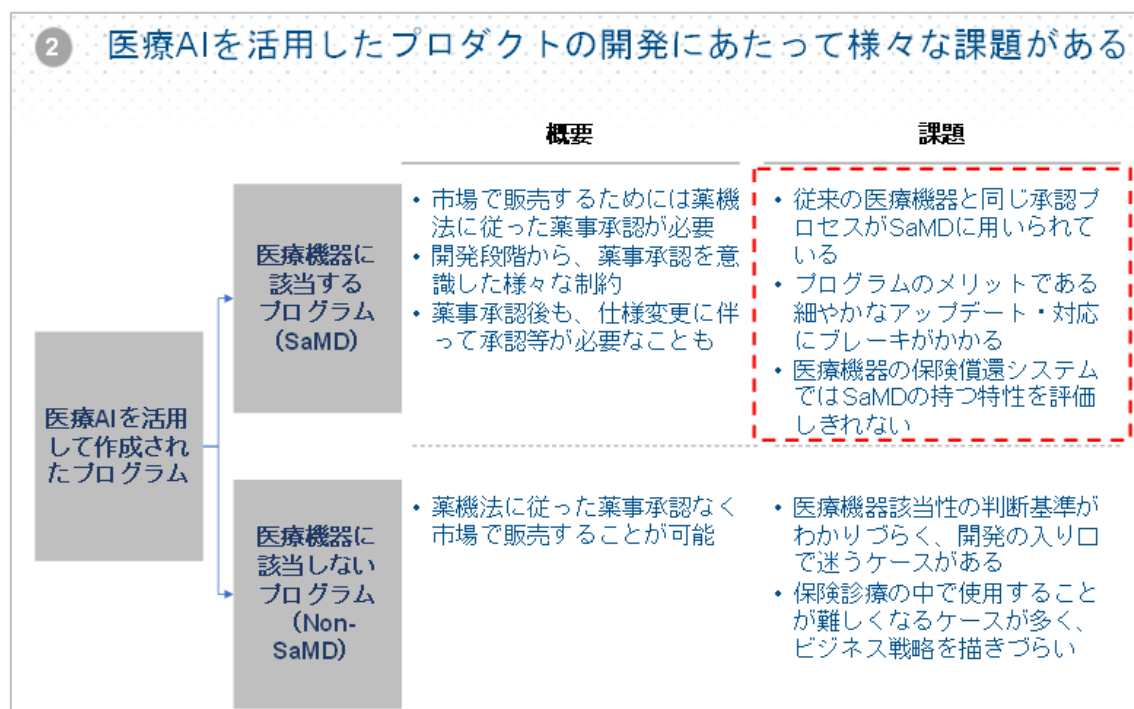
【原聖吾氏】 この観点で見たときには、国内で様々な種類がありますが、日本ではごく一部しか保険の加算が取られていません。実際に出口として保険の償還を得るような医療プログラムは限られています。



【原聖吾氏】また、この SaMD は、国内で認可を受けているものは全体で 300 種類ほどありますが、治療用に使われるものはごく一部に限られています。



【原聖吾氏】 翻って他国を見ると、FDA では 343 品目におよぶ AI あるいはマシンラーニングを活用した医療機器のリストを公開していますが、日本では今これが実際のぐらいの数になっているのかも確認できません。このように、AI を活用した製品、サービスというものが見えにくくなっているという状況です。



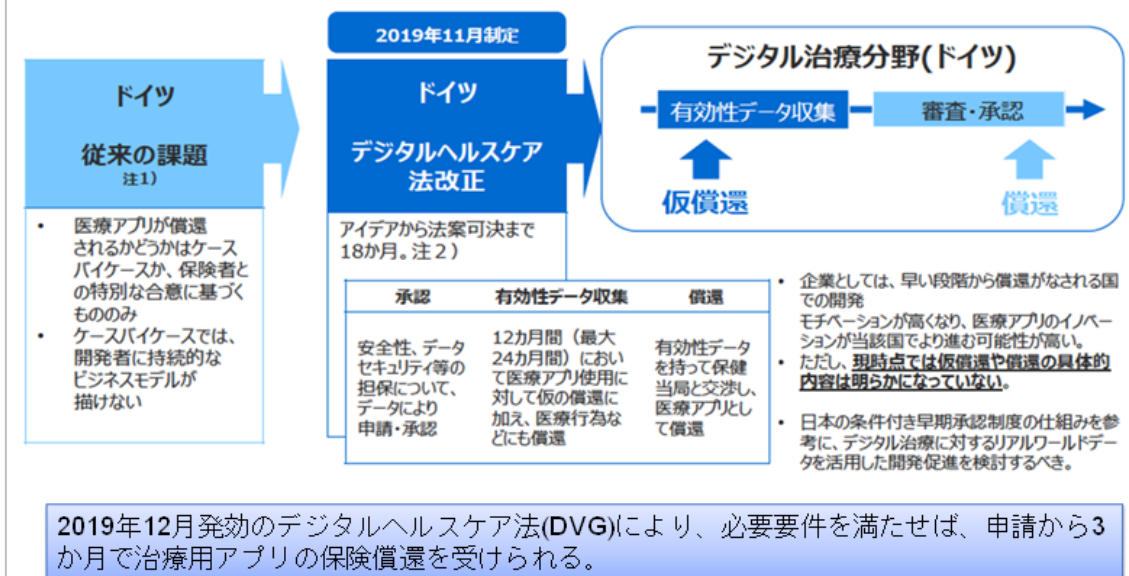
【原聖吾氏】 医療 AI を活用したプロダクトの開発に当たって、課題があろうかと思います。従来の医療機器と同じような承認プロセスが使われている。アップデートに対応していない。あるいは保険償還システムが SaMD に対応していない。このあたりが大きな課題と考えられます。

② 各国でAIを含めたデジタルヘルスに対する取組が強化されており、日本でも検討開始されているものの、詳細比較と検討が必要

	日本	韓国	米国	ドイツ
政策名	DASH for SaMD（2021年4月より検討開始）	政策名不明だが、デジタルヘルス製品の開発強化 医療機器全体の規制緩和	Digital Health Innovation Action Plan	Digital Healthcare Act
開発関連の施策	- SaMD審査室の設置 - 窓口の一元化 - 審査制度の検討 - 保険収載における考え方の検討	具体的な内容に踏み込んだAI医療機器開発・承認審査基準ガイドライン（英語版あり）	開発関連のガイドラインが多数発出されている	デジタルヘルスに特化したベンチャー支援の実施
薬事承認	- 現時点では、デジタルヘルスに特化した薬事承認スキームはなく、従来の医療機器と同じ承認スキームで評価される - IDATEN制度	AI医療機器のみに特化した承認迅速化を行っているわけではないが、元々医療機器産業規制緩和が取り組まれており、審査期間の短縮や手数料の値下げなどが行われている	- Precertification Program - Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan	- MDR（EU圏内の医療機器認可）を取得することが条件 - 安全性・データセキュリティ・品質の規格を満たせば薬事承認（有効性は評価しない）
保険償還	なし	なし	なし	暫定的に保険償還し、上市後の一年間で有効性データを収集
保険償還	- 保険償還で加算を得られた製品は数えられる程度（欠項） - 今年中に大まかな方針を決める予定	不明	上市後に収集されるデータを安全性・有効性評価に活用し、保険償還を目指す	市販後に収集された有効性データを基に交渉により償還価格を決定する

【原聖吾氏】他国との比較ですが、他国では進んでいるような事例も多く見られます。特に、ドイツがデジタルヘルスに特化した制度を作ってしまして、現在、世界的にもSaMD、デジタルヘルス、AI を活用した出口として、多くの上市が進んでいる国になっています。

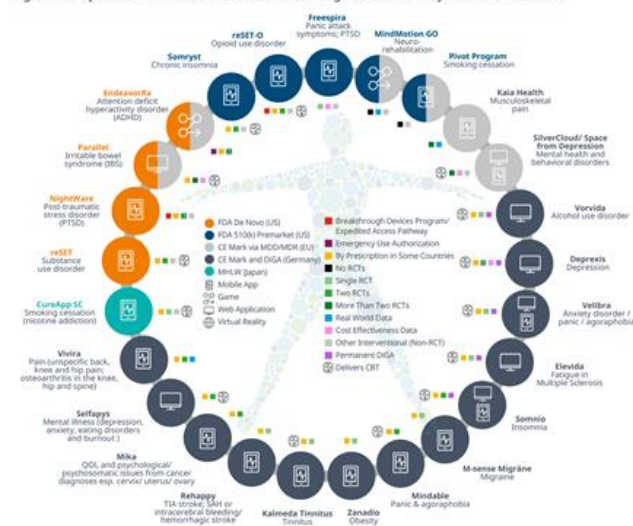
2 ドイツではデジタルヘルスケア法を2019年に制定し、特に治療用アプリについては開発加速化に向けたスキームを導入している



【原聖吾氏】ドイツは、デジタルヘルスケア法を2019年に制定し、その中で、早いタイミングで、安全性が担保された段階で承認をします。その後、仮償還をし、有効性データを収集して、保険償還をしていくということで、早いタイミングで承認・保険償還をする仕組みを整備しています。

2 デジタルヘルスケア法を制定したドイツは、デジタル療法の承認数で日本に大幅に先行

Digital Therapeutics with Market Authorization through Select Pathways and their Features



各国で承認済みのデジタル治療製品数
(2021年6月時点)



Source: IQVIA Digital Solutions Database, Jun 2021; IQVIA Institute, Jun 2021

【原聖吾氏】結果的に、デジタル治療の製品の数は、日本では限られています。ドイツでは13品目ほどということで、この領域でも先行しているような事例になっています。

3 開発されたAIプロダクトが臨床現場に浸透し真のDXにつなげるためには、補助の拡充や診療報酬上の評価が必要

	事例	課題
a 業務フロー	- 問診 - 支払 - 物資の在庫管理 - 医療文書作成 等	- 医療資源の限定的な医療機関ほど初期導入費用・継続費用の捻出が困難 - 大規模医療機関では、臨床現場のニーズが必ずしも経営の意思決定に反映されない
b 医療行為 (予防・診断・治療)	- PHR - 画像診断 - 手術支援 - 疾患治療システム - 遠隔診療 等	- 事業者にとってはAI医療機器が従来医療機器と同様の評価方法にあり、開発に時間を要するため、事業戦略を見込づら - 医師からすると、責任範囲が明確でない
c 創業等	- 医薬品開発 - ゲノム医療 - 基礎研究 等	AIの提案を医師の判断にどう反映させるかが、明確になっていない(スクリーニング、診断、ダブルチェック)

臨床現場が必要とするプロダクトの導入に必要な費用をカバーできる補助の拡充

臨床現場を支援するAIプロダクトに対して導入促進のため診療報酬上明確かつ十分に評価

【原聖吾氏】最後に、臨床現場への浸透における課題についてです。実際に開発された AI のプロダクトを臨床現場に浸透させるという観点では、課題があります。特に、大きな課題になってくるのは、医療現場の中で、導入のための費用の部分です。また、医師からすると、この責任範囲が明確になっていないところです。このあたりが大きな課題であり、この浸透に当たってはハードルがあるということが課題として挙げられています。

まとめ

- 医療分野におけるAI活用においては、データの蓄積・AIプロダクトの開発・現場への浸透と、それぞれのステップで課題がある
- 医療データの蓄積・連携においては、そもそもデジタル化されていない情報が多いことに加え、データの精度のばらつき、活用用途の限定等が課題となっている
- 医療AIプロダクトの開発においては、特に医療機器に該当するプログラムの承認・アップデート・保険償還の課題がある
- 現場での活用にあたっては、予算補助の充実と診療報酬上の評価が必要

【原聖吾氏】データの蓄積、開発、現場への浸透という三つの部分、それぞれのステップに課題があって、課題それぞれを解決していくことが、今後、医療 AI プロダクトを活用していく上では重要であろうと考えています。

【以下、児玉安司】 AI からプロダクトを作っていくということでも、日本は立ち遅れている中で、一定の範囲で SaMD の製造販売が行われています。しかし、SaMD として提供することと、AI のまま提供していくということには、実際には大きな差があります。

3. SaMD から医療生成 AI へ

3 SaMDから医療生成AIへ

American Medical Association Augmented intelligence in medicine

UPDATED APR 5, 2024.

CONTENTS

[Artificial intelligence vs. augmented intelligence](#) |
[Principles for AI development, deployment and use](#) | [Physician sentiments on AI](#) |
[Featured updates](#) | [CPT® and AI](#) | [Latest AI news](#) | [AI and practice management](#) |
[AMA Board reports and policy](#) | [AI learning on Ed Hub and JAMA Network™](#)

Artificial intelligence vs. augmented intelligence

The AMA House of Delegates uses the term augmented intelligence (AI) as a conceptualization of artificial intelligence that focuses on AI's assistive role, emphasizing that its design enhances human intelligence rather than replaces it.

Principles for AI development, deployment and use

The AMA is committed to ensuring that AI can meet its full potential to advance clinical care and improve clinician well-being. As the number of AI-enabled health care tools continue to grow, it is critical they are designed, developed and deployed in a manner that is ethical, equitable and responsible. The use of AI in health care must be transparent to both physicians and patients.

これは今年の 4 月に American Medical Association (AMA) が出した「Augmented intelligence in medicine」というコメントです。AMA は長らく、AI というのは augmented intelligence、つまり医師の力を増強していく intelligence であって、医師が使う医師と協働していくツールなのだという位置付けを明瞭にしています。

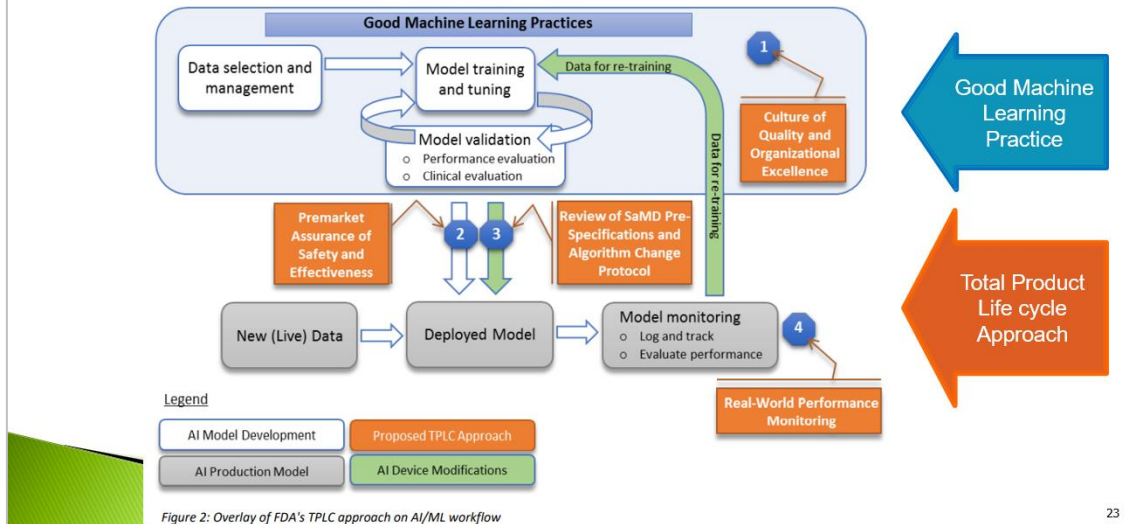
SaMDからOpen AIへ

	SaMD	Open AI
著作権	○	Prompt制御など
個人情報保護	○	Prompt制御など
製造物責任	○	ハルシネーションなど
薬機法の承認	手続整備済	？

SaMD の場合は、著作権の問題はそれほど大きくないし、個人情報保護、製造物責任についても、どのぐらいの正診率でどのぐらい外すのかというようなことも、説明も warning もできるわけです。現在、SaMD については、薬機法の承認についても、手続が整備されています。

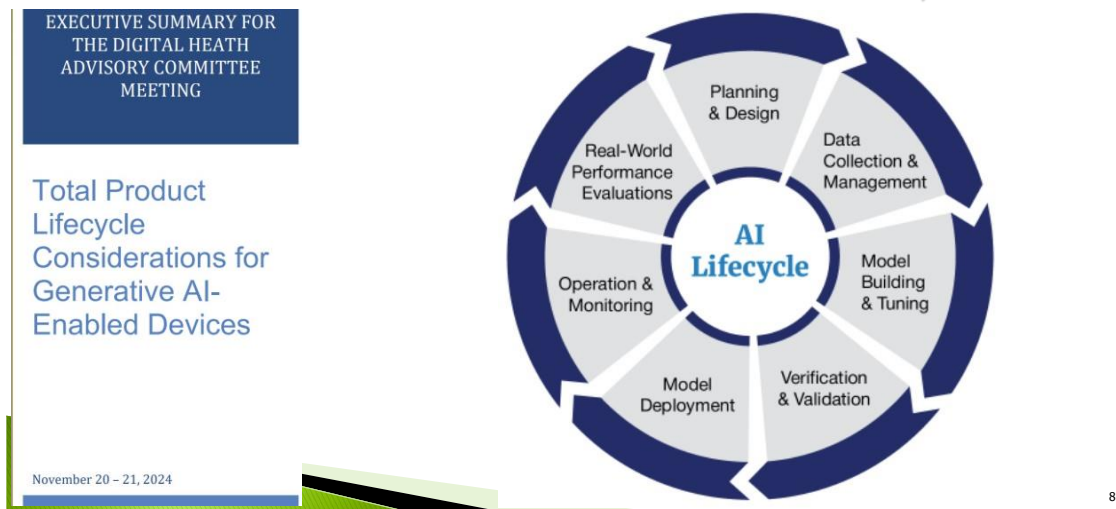
一方、Open AI については、著作権についてもどのような制御をかけるか、個人情報保護にどう制御をかけるか。AI が現場の情報を取り入れて育っていったときにもハルシネーションなどの問題が起こります。情報にバイアスもかかります。例えば、現場で特殊な症例を経験してしまった AI が、全体の確率に関係なく違う答えを出し始めるというようなバイアスがかかることは十分にあり得ることです。そういったものをどう制御するか。生成 AI を医療現場で使い、現場の情報でさらに発展生成させていくことをどう扱うかについては、まだ薬機法の準備が十分できていません。

FDA: Artificial Intelligence and Machine Learning in Software as a Medical Device September 21, 2021



アメリカのFDAは、元々のプロトコルを作り、アルゴリズムを作り、そしてSaMDにしていくところを、Good Machine Learning Practice といっていますが、これは、Good Clinical Practice など、医薬品・医療機器の承認関連において、いつも出てくる言葉と同じパターンで、machine learningをやるのであれば、そのプロセスを適切に管理しようという考え方です。それに加えて、Total Product Life Cycle Approach、つまり現場に出たAIが育っていく、それをできるだけオンタイムに捉えながら承認していくようなアプローチを取れないか検討していると報じられています。

FDA: Total Product Lifecycle Considerations for Generative AI-Enabled Devices (November 20-21, 2024)



さらに FDA は、2024 年 11 月 20 日から 21 日の Advisory Committee Meeting に発表したステイトメントで、医療生成 AI について、計画立案からリアルワールドでの品質評価までの AI ライフサイクルを、プランニングと設計、データの収集と管理、モデルの構築とチューニング、検証と妥当性確認、モデルの展開、運用と監視の一連のサイクルとして捉え、さらに充実させていくことを表明しています。欧米も中国もますます加速しています。

ICMRA horizon scanning assessment report 6 August 2021 EMA ビッグデータ・タスクフォース

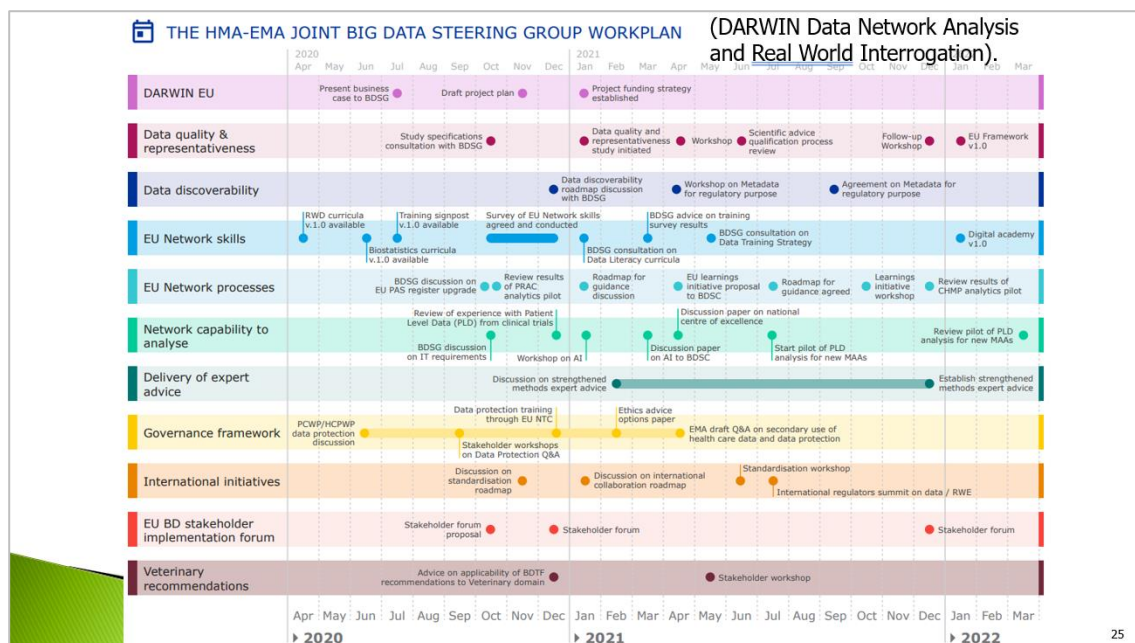
- **ビッグデータ・トレーニングカリキュラムと戦略の立案**

(Develop a Big Data training curriculum and strategy based on a skills analysis across the Network, collaborate with external experts including academia, and target recruitment of data scientists, omics specialists, biostatisticians, epidemiologists, and experts in advanced analytics and AI)

- **AIの検証を行うネットワークの能力の強化**

(Strengthen the Network ability to validate AI algorithms)

ヨーロッパでは、EMA、ビッグデータ・タスクフォースというものがあり、AI の検証を行うネットワークの能力の強化が必要などと言われています。



さらに AI を作り、使い、育てるために、DARWIN と呼んでいるデータ基盤の生成を、2022 年の春ごろまでには終えて、いよいよ AI がどんどん出てくるので、今回、5 月 22 日の AI Act の制定につながっているということです。

[個人情報保護委員会](#) > [広報・お知らせ](#) > [注意情報](#) > 生成AIサービスの利用に関する注意喚起等について

生成AIサービスの利用に関する注意喚起等について

生成AIサービスの利用に関する注意喚起等について

我が国において、現在、生成AIサービス（質問・作業指示（プロンプト入力）等に応じて文章・画像等を生成するAIを利用したサービス）が普及していることを踏まえ、当委員会として、別添1のとおり、生成AIサービスの利用に関する注意喚起等を行うこととしました。

なお、生成AIサービスであるChatGPTを開発・提供するOpenAI, L.L.C.及びOpenAI OpCo, LLCに対しては、別添2に記載の概要のとおり、注意喚起を行いました。

■ [【別添1】生成AIサービスの利用に関する注意喚起等 \(PDF : 286KB\)](#)

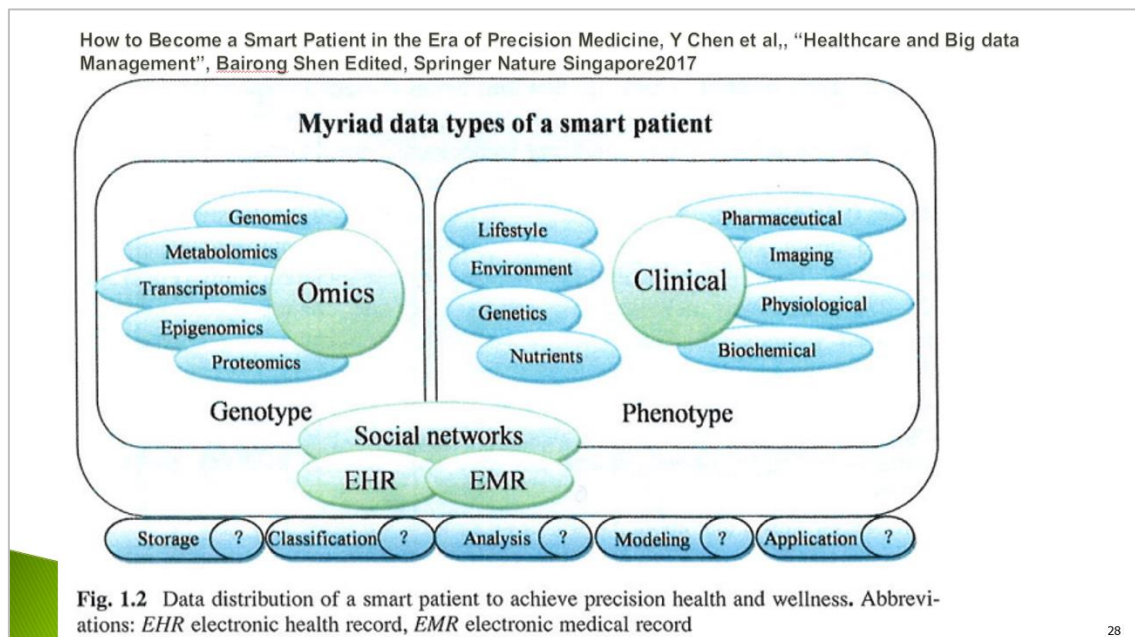
■ [【別添2】OpenAIに対する注意喚起の概要 \(PDF : 180KB\)](#)

日本の現状は、生成 AI サービスの利用に関する注意喚起ということで、個人情報保護委員会の注意喚起が物議をいろいろ醸している状況です。

個人情報保護委員会は、2025 年 2 月 5 日に、「個人情報保護法の制度的課題に対する考え方について（個人データ等の取扱いにおける本人関与に係る規律の在り方）」と題する文書を発出し、「統計情報等の作成（注 1）のために複数の事業者が持つデータを共有し横断的に解析するニーズが高まっていること、特定の個人との対応関係が排斥された統計情報等の作成や利用はこれによって個人の権利利益を侵害するおそれが少ないものであることから、このような統計情報等の作成にのみ利用されることが担保されていること等（注 2）を条件に、本人同意なき個人データ等の第三者提供及び公開されている要配慮個人情報の取得を可能としてはどうか（注 3）。

注 1：統計作成等であると整理できる AI 開発等を含む。（後略）」として、注目を集めました。今後、我が国において、A I 開発が一気に加速することも期待されます。

まとめー「賢い患者」 Smart Patient という夢



蘇州大学のチェン（Chen）教授が書かれた「How to Become a Smart Patient in the Era of Precision Medicine」という論文で、台湾の東呉大学のバイロン・シェン（Bairong Shen）教授が編集した『Healthcare and Big data Management』という本の巻頭論文として掲載されました。2017年にシンガポールで出版されました。

加加速度的に増えていく膨大な量の個人データがある、ゲノムのデータもある。ゲノムからどのように protein になっていくかという過程全部の、Omics data と呼ばれるものもある、臨床データもある。また、その人がスマートウォッチを装着していれば、その人自身の生活スタイルや睡眠時間などいろいろなことも分かるし、家族のゲノムデータもある。Omics も Clinical も、様々なデータをソーシャルネットワークでできるだけ統合して、EHR（electric health record）、EMR（electric medical record）を大きく広げて構想していこうという提案です。

How to Become a Smart Patient in the Era of Precision Medicine, Y Chen et al, Bairong Shen Edited, "Healthcare and Big data Management", Springer 2017

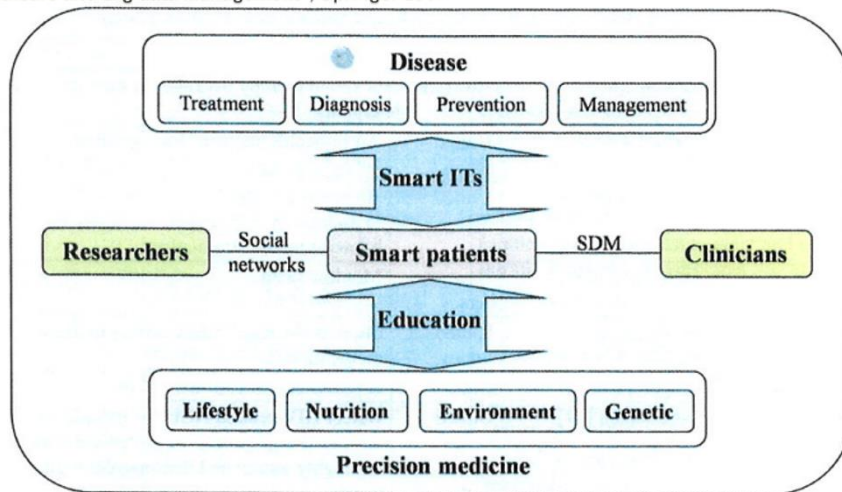


Fig. 1.3 Participatory model of a smart patient. Abbreviation: *SDM* shared decision-making

Smart patient のところには病気の情報も生活の情報も皆集まってくるわけですから、その情報を臨床医（clinicians）と共有して、SDM（shared decision making）を行っていく。あるいはその情報をソーシャルネットワークの中で、研究者（researchers）が利用していく。

○科学・産業政策と国際競争を考慮した**合理的政策の充実**

○国民へのさらなる情報提供による**合理的期待の形成**

○**個人情報**の適正かつ効果的な活用が**新たな産業の創出**並びに**活力ある経済社会**及び**豊かな国民生活の実現**に資する（個人情報の保護に関する法律第一条）

研究規制であれ個人情報保護規制であれ、科学・産業政策と国際競争を考慮した合理的政策を形成し、国民の側に、医療 AI についての合理的な期待の形成を図っていく必要があります。このことが、「個人情報保護法第 1 条に定める個人情報の適正かつ効果的な活用が新たな産業の創出並びに活力ある経済社会及び豊かな国民生活の実現に資する」という個人情報保護法そのものの実現に資するものと考えます。

テーマ4 医療データ基盤の構築と運営手法の検討

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

研究開発責任者

合田 憲人 情報・システム研究機構 戦略企画本部 ・ 教授

同席者

武田 憲彦 東京大学 医学部附属病院 ・ 教授

【合田憲人教授】 タイトルは「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」というものです。

要旨: 研究開発の概要

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

解決すべき課題

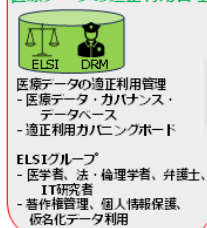
1. 大規模医療データ基盤

- 1-1 大規模医療データの収集・管理が困難
- 1-2 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

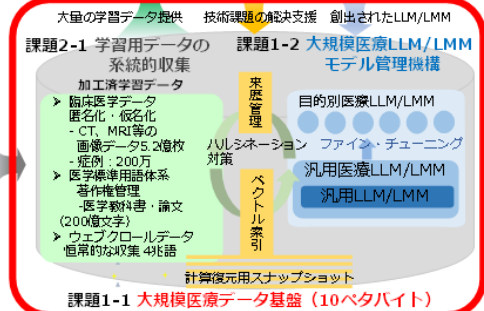
2. データ収集・持続的な利活用

- 2-1 医療 LLM/LMM 開発のための系統的なデータ収集が困難
- 2-2 医療データの適正利用のための仕組みがない
- 2-3 法・倫理的課題の解決が必須

課題2-2, 2-3 医療データの適正利用管理



セキュリティ



医療データ・医療LLM/LMMの利活用を促進する大規模医療データ基盤

このスライドは全体のオーバービューで、医療 LLM/LMM の研究開発を促進するため、まず医療データを集めて、加工して蓄積して、利用可能とする医療データ基盤というのを作るのが一つ目の目標になります。

次に、この医療データ基盤上のデータを、安全・適正に利活用できるようにするための仕組みを作るのが二つ目の目標になります。

通常、医療データ基盤というと、LLM/LMM の学習に必要な医療データをためて、管理して使えるようにするというのが一番コアな役割になります。そして、医療データ基盤にとどまらず、新しいチャレンジを入れることで、医療 LLM/LMM の研究開発に本当に資するものを作ることと、医療データの利活用の推進を目指しています。

これから創出する予定の医療データ基盤は実態は計算機が並んでいるものです。そこにデータをためたり取り出したりするソフトウェアが動いていて、様々なデータを検索するデータベースが動いています。緑色の部分が、コアとなるデータの収集・蓄積です。医学の教科書とか論文のようなテキストのデータをコンテンツ利用契約を締結して読み込ませたり、匿名化・仮名化等が施された CT、MRI 等の画像のデータを

医学系の学術研究機関から集めたりして、このデータセットを LLM/LMM の研究開発に利用できるようにするというのが一番の役割、2-1 になります。

LLM/LMM の機械学習を行わせるために GPU やクラウドのシステムを準備していて、様々な LLM/LMM のモデルができています。

LLM/LMM というのは一つではなく、最終的には医療目的ごとに特化した生成 AI ができ上がってきます。多種類のモデルが何世代にもわたって出てくるため、それらを適切に蓄積し、管理しないと、使う側が使いきれないという問題が出てきます。

そこで、医療データ基盤では、こういった生成されたモデル、保存して管理するような大規模医療 LLM/LMM モデル管理機構というものも新しく作ることを構想しています。

また、LLM では情報空間に存在しないような誤情報・偽情報を生成してしまうハルシネーションが課題になっています。その防止のための研究開発を支援する機構に関する研究開発を行います。この黄色い部分の技術課題の解決支援を行っていくことを計画しています。

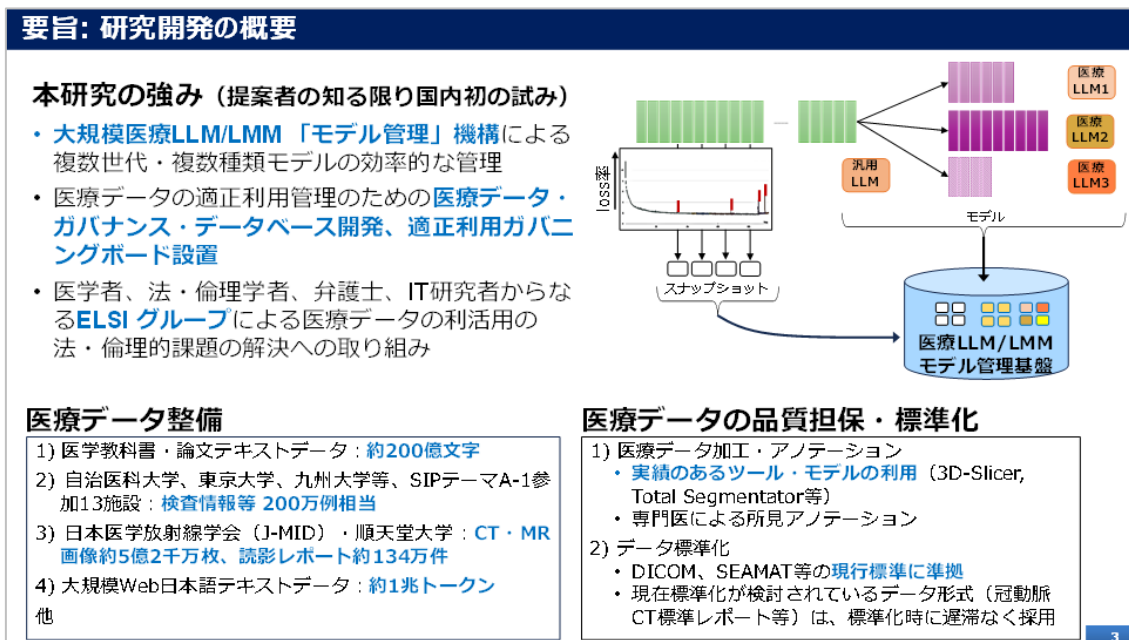
ここまでの医療データ基盤を中心にした話です。もう一つの、データを安全かつ適正に利活用するための仕組みが、左側の緑の部分です。ELSI の課題は主としてこちら関係がありますが、ここでは二つのことをやろうとしています。

一つは、医療データの適正利用管理です。ここで集めてくる医療データというのは、利用条件がそれぞれ異なり、複雑な条件が課されています。個人情報保護の観点からの条件もあれば、著作権との関連で権利侵害とならないようにするための条件もあります。研究者はこのような個々のデータの条件を、紙の上で管理して、これは誰が使っていていい、これは自分が使えるかというのを判断していますが、ここをシステムティックに動くようにしようということで、データごとの要件をまとめたデータベース、

医療データ・ガバナンス・データベースというのを新しく開発する予定です。利用者がこのデータベースにアクセスして自分の情報を入れると、自分が使えるデータが分かるというイメージです。

さらに、このデータベースというのは、作るだけでは駄目で、適正に運用されないと間違いが起きてしまうので、このガバナンス・データベースが適切に運用されているかを管理する、適正利用ガバニングボードという委員会を設けて、ここで適切に運用を見ながら適正利用管理ができるようにするということを目指しています。

最後は ELSI のグループです。右側のモデルの生成においてもそうですが、様々な法的、倫理的課題が出てきますので、ここにあるような医学者、法・倫理学者、弁護士、IT 研究者などからなる ELSI グループというのを設けて、こういった研究を進める上での課題については、常にここで検討しながら進めていくという体制を取るということを考えています。



研究の強み、新しいチャレンジというのは、データをためるのは当たり前なので、それに加えて、LLM/LMM のモデルを管理する仕組みを作りましょうという話と、

医療データ・ガバナンス・データベースというものを作って、データの適正利用をできるようにしましょうという話と、最後、ELSI グループの話になります。

一つ目のモデル管理は、単にデータを管理するだけではなく、LLM に特化した話ですが、複雑・重要な課題です。実際に医療 LLM ができるまでの流れを模式的に表したものです。学習データを読み込んでいって学習していくと、このグラフが下に行くほど精度が上がっていき、モデルが完成するわけです。

最初から医療のデータを読み込ませてモデルを作るのではなく、最初は汎用の言語のデータを流し込んで、汎用 LLM というものを作ります。これをベースとして、これに追加で、医療の様々な分野とか、循環器のデータですとか、そういったものを新しく追加で学習させることによって、様々な用途に応じた医療の LLM ができるという流れになります。従いまして、このようなモデルがたくさんできてくる、かつ、世代ごとにできてくるため、まずこのモデルを管理しましょうというのが、このモデル管理基盤の役割になります。

もう一つは、汎用 LLM を作る過程において、学習すればするほど精度は下がっていくというのが一般的なのですが、突然、精度が落ちてしまい使えなくなってしまうことが起きて、また一から学習をやり直さなければいけません。それを防ぐために、定期的に途中経過として、スナップショットというのを保存しています。スナップショットを適宜撮って、適切に管理して、何か起きたときは直ぐに戻れるようにしておくというのも、このモデル管理基盤の役割になります。

続いて、適正利用と ELSI グループについて説明します。

まず、医学の教科書・論文のテキストデータは、購入し、オープンでアクセスできるものは無料で持ってきて、この緑の部分のところだと紫の部分の学習に使います。加えて、いくつかの大学病院の中から検査情報などが集められます。

NII と日本放射線学会とが連携した J-MID というプロジェクトで CT や MRI の画像を毎日収集するというプロジェクトをしています。これがいま 5 億枚以上ありますが、このようなデータやそれに付随する読影レポートも一緒に集めて、今回のプロジェクトに使われているようになります。同時に、Web 日本語テキストデータ、要は Web で検索して行って、そこから機械的に様々な日本語の文章を集めるということをしなが
ら、こういったものも機械学習に使っていくことを計画しています。

評価基準 1: SIP の意義の重要性や趣旨に合致していること。

基礎研究から社会実装までの一貫通貫の研究開発

- ・ 医療データ基盤に関わる基礎研究（医療 LLM/LMM モデル管理・技術課題解決支援）
- ・ 適正に利用管理された医療データ利活用の促進に至る社会実装

医療データ利活用の社会的受容性を高める社会実装に向けた戦略

- ・ ELSI グループが法的・倫理的課題に精力的に取り組み、医療データ利用のメリット・価値を認める社会の合意形成へ寄与



社会的課題の解決や日本経済・産業競争力の強化

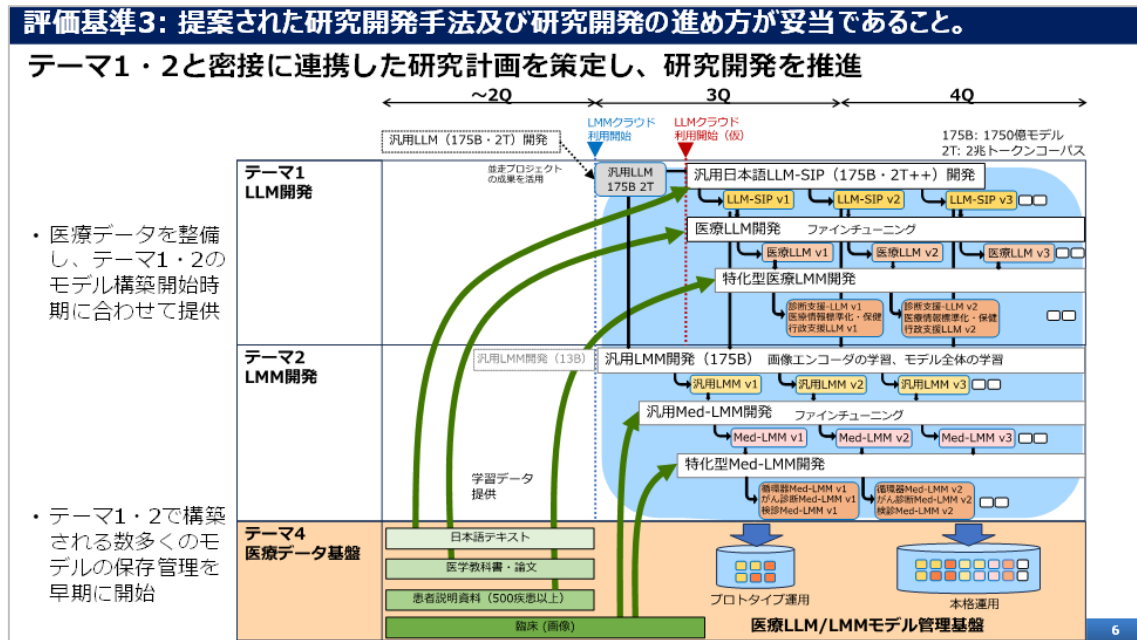
- ・ 医療データの利活用を進めるための社会的課題（利用権、法・倫理課題の複雑さから医療データの利活用が進まない）の解決に向けた貢献
- ・ 医療分野での LLM/LMM 開発・利活用促進による経済・産業競争力の強化に向けた貢献

SIP の中の意義としていくつか定められている点がありまして、それらにこのプロジェクトがどう対応していくかを表したスライドです。

SIP ですので、基礎研究だけではなくて社会実装までやることが重要です。我々の研究では、ハルシネーションといった基礎研究への支援だけではなく、それを活用すると同時に、医療データ適正利用を実現することによって、社会実装にも貢献するというを考えています。また、医療データ利活用の社会的受容性を高めるという点については、課題に取り組むことによってこれを進めようと考えています。

このような取り込みを進めることによって、医療データの利活用を進めるという社会的課題、要は様々な法的なところからデータの利活用が進まないことの解決に向け

た貢献をしたいと考えていて、こういった LLM/LMM が医療分野で利活用されることによって、わが国の経済・産業競争力の強化にもつながるだろうと考えています。



これは非常に細かいのですが、データを用意して、テーマ1や2のLLM/LMM開発に資するということで、一緒に連携してやっていくため、それに向けたマイルストーンを描いたものです。実はLLMとLMM、既にこのテーマは研究が始まっています。最初に汎用LLMという日本語のLLMのモデルを作って、全てここを起点にして、医療上のLMMを作ったりしていくことになります。

この汎用LLMというのは、いわゆるChatGPTの日本語版のようなものでして、世の中の日本語のテキストをたくさん学習して作ったモデルです。これはNIIの黒橋所長の下で、LLM-jpというプロジェクトがあり、そこでいま作っています。我々のこのSIPのプロジェクトで使えるモデルが、この3Q、9月の終わりにできてきます。

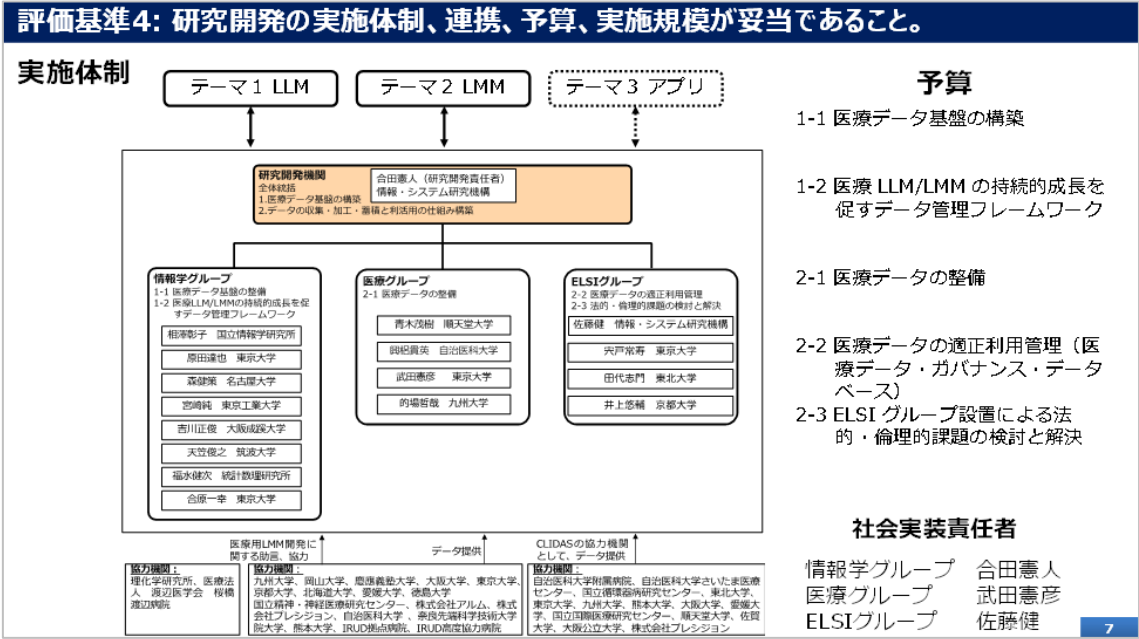
これができると、今度はこれに対して医療関係のテキストを読み込ませて、機械学習させていくことで、医療用のLLMを作ります。ここから、特定の病気の診断支援や

医療情報標準化、目的に応じたデータを学習させることで、特化型の LLM ができるという流れになります。

実は LMM についても、画像だけではなく、テキストと画像の両方を使うということになるため、スタートはテキストの LLM のモデルから始まります。これを使って、これに対して画像のデータを新しく追加で学習させることによって、様々な目的に応じた医療の LMM ができるという流れになります。

こういった流れに対して、まず一つ目は、データを提供します。医療データ基盤の側で、日本語テキストから、適切な個人情報保護が行なわれた画像等の臨床情報を集めることが我々のミッションになっています。そのため、こういったデータを整理して、かつ、この計算が始まるタイミングで提供するということに向け、準備を進めているところです。

同時に、実際にこの学習が始まると、この第三四半期、第四四半期の間にモデルのデータが出てきますので、そのモデルのデータをモデル管理基盤で蓄積できるようにすることが、ここでの我々の役割になります。



体制です。ROIS が取りまとめを行い、15 グループあると思いますが、それぞれ情報学と医療と ELSI のグループに分かれて、ここに挙げられた方々が共同研究者として参画して、研究を進めていく計画としています。

情報学のグループは、データの基盤ですとか、様々なモデル管理基盤といった部分の研究開発を行います。医療グループの方々には、主に医療データの整備を行っていて、ELSI のグループでは、主にその医療データの適正利用管理と法倫理課題の検討を行っていきます。

また、こういった取り組みをすることによって、テーマ 1、2、3 とともに密に連携します。テーマ 3 はいくつかアプリケーションが出てきて、既に話を進めています。ここでの課題を解決するために、我々の医療データ基盤の機能を使うことについて検討しているところです。

評価基準5: 提案されたアウトプットとしての技術が優位であること。

技術的優位性

- ・ **10PB大容量ストレージ**が、大量のデータ（医療教科書・論文約 200 億文字、臨床データ約 200 万例、医療画像情報約 5.2 億枚等）を**SINET高速ネットワーク（400Gbps）**とも連携して収集、蓄積
- ・ 国立情報学研究所の**高度なセキュリティ技術を用いた最新の攻撃検知システム**を拡張し、サイバー攻撃被害を最大限に抑止
- ・ 医療用仮想プライベート・クラウド
- ・ **ハルシネーション対策**、アンラーニングほかの技術課題解決を支援

提案者らの実績

- ・ 最先端のクラウド、データベース、ネットワーク、セキュリティ等のIT研究に従事するとともに、最先端学術情報基盤サービスを全国の大学等に展開
 - ・ 文部科学省「**研究データの活用・流通・管理を促進する次世代学術研究プラットフォーム**」
- ・ 医療ビッグデータクラウド基盤の運用知見を活用
 - ・ **2017年より6つの医療系学会から5億枚以上の医療画像データを収集、42件の医療画像解析AIを開発**

テーマ1・2・4協働によるハルシネーション対策

このスライドは情報学的な技術の話になります。医療データもそうですし、モデルのデータも非常に巨大なので、10PB（ペタバイト）という単位の大容量のストレージを用意し、かつ、SINET、これは NII が運用している学術情報ネットワークですが、

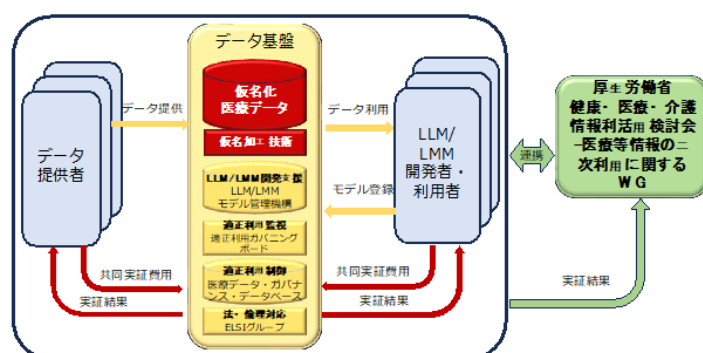
大学や研究機関は、実は普通のインターネットと違って専用のネットワークである SINET でつながっています。ものすごく速い、全国 400Gbps でつながるもので、そこで高速にデータを集めたり取り出したりすることを実現します。また、この種のものではセキュリティが最重要課題ですので、セキュリティに対しても非常に高度なものを入れながら構築していく予定です。

NII では、NII-SOCS といまして、大学に向かって流れるトラフィックの、中身は見ませんが、通信のパターンを監視する仕組みがあって、その中で、これは明らかにセキュリティの攻撃であるということが分かると、それを大学に通知するというサービスを行っています。ものすごく膨大な量のトラフィックのパターンを常に毎日解析しながら発見するのですが、今回、そのグループとも連携しまして、最新の攻撃検知システムを入れて、万一、我々医療データ基盤に攻撃があった場合でも、すぐにそれを抑止できるような仕組みを入れようと考えています。

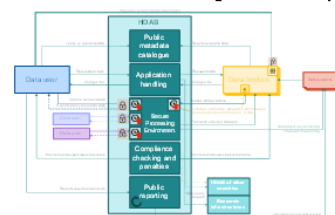
評価基準6: 提案された出口戦略が優れていること。

「仮名化」した医療データをLLM/LMMに活用する実証実験を計画

- pseudonymisation (GDPR Article 4(5))
- EU/UKでもまだ検討中の問題
- 厚生労働省の医療等情報の二次利用に関するWGと連携
- データ提供者・利用者からの「共同実証費用」により運営
 - ・ 現行法の下で、共同研究契約、データの委託契約により実施



EUにおけるデータ基盤
European Health Data Space (EHDS)
(2024/4/24 EU会議でEHDS Regulationを承認)



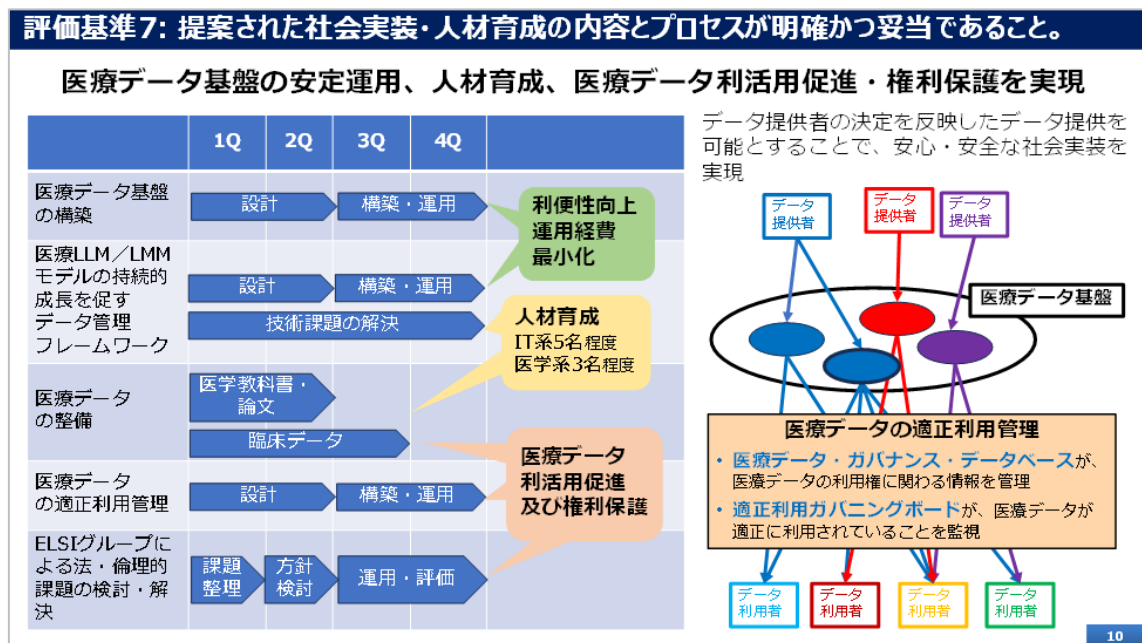
UKにおけるデータ基盤
Trusted Research Environment (TRE)



ここからが、二つ目のデータ利活用、適正利用管理に関する話です。ここで我々が目指しているのは、仮名化した医療データを適切にデータ提供者と利用者の間で流通

させるような仕組みを作るのが一つの目標になってきます。これについては、EUにおいても EHDS が最近始まったばかりで、UK では TRE が始まったところで、まだ海外でも検討が始まったところです。そのため、こういったものを勉強しながら、これはまだ検討中で最終形ではないですが、医療データ基盤をハブとして、データを提供する側と利用する側が、この利用する側には LLM や LMM の開発者も含まれますが、データをうまく利活用できる仕組みを作っていこうと考えています。

また、こういった仕組みを回すときは、医療データ基盤のシステムの運用費も含めて、ビジネスモデルを作る必要があるため、費用についてもうまく流れる仕組みを検討していきたいと考えていますし、厚労省の医療等情報の二次利用に関するワーキンググループの動きも見ながら、連携しながら進めていこうと考えています。



これは、全体のスケジュール感です。最終的には社会実装が非常に重要で、単にものを作って研究して終わりではなくて、その先も見据えた形で研究を進めていく予定です。また基盤のシステムについては、利便性の向上を図るとともに、運用経費も最

小化しないと続かないため、そういった部分を技術的な観点から行っていきたいと考えています。

医療データ利活用については、データの提供者が上にいて利用者がいるとすると、提供するデータには様々な制約があって、その制約を満たした適正な利用を管理するのが目的になります。ですので、医療データ・ガバナンス・データベースを作って、それを適正に運用する仕組みを作ること、利用者が自分の必要なデータを必要な時に適切に活用できるような仕組みを作っていくことで、医療データの利活用促進と、元々ある著作権の権利保護の両立を図っていきたいと考えています。

評価基準8: 産業界(民間企業)等からの貢献(人的、物的投資を含む。)の見込みが大きいこと。

本事業終了後の医療データの利活用

- 医療データ基盤に整備された大規模医療データは、本事業終了後においても、医療機関、研究機関、企業等による利活用の継続・拡大が期待される



医療データ活用の推進

- IBM 社、プレジジョン社、アルム社が、来年度以降、共同実証費用を支払う検討を進めている
- 今後、更なる民間企業、医療機関、研究機関、行政機関、等の共同実証の有償参画を期待

テーマ1, テーマ2 の社会実装への貢献

- 医療データの収集とその適正利用管理の機能（データに関する患者の同意状況、許諾された利用目的、利用許諾者、利用期間、データ加工種別、更新情報等に基づくアクセス制御）を継続的に提供し、テーマ1, 2 の社会実装に貢献する

補正予算で、本年度開始のプロジェクトであり、このような仕組みを作ること、この事業終了後も、基盤とデータの利活用は継続、拡大が期待されます。そのため、拡大させないといけないわけですが、それに向けて、今後の有償利用等も含めたビジネスモデルの確立ですとか、あとはテーマ1や2、LLMやLMMの社会実装にも貢献していきたいと考えています。