FG AI4H会合報告

慶応義塾大学大学院　政策・メディア研究科　川森（かわもり）　雅仁（まさひと）

1. はじめに

ITUとWHOが共同で開催しているFocus Group on AI for Health（FG-AI4H）は、コロナ禍で、オンライン開催となり、2022年2月15日から17日にN会合が開催された。今回で14回目の会合となる。

2. 出力文書の進捗

以下に今回の会合でアップデートされた出力文書の概観を述べる。

2.1. DEL 2-AI4H規制のベストプラクティス

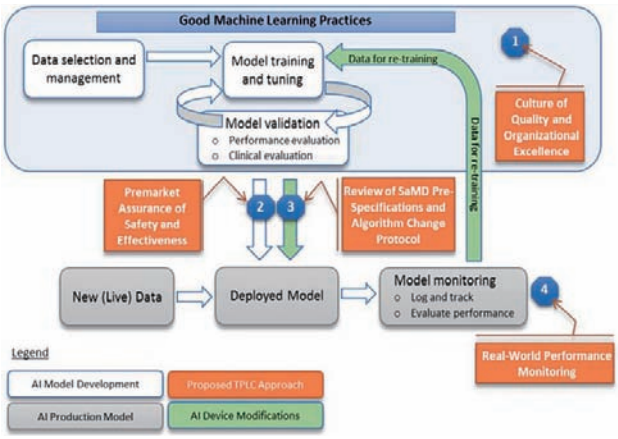
本文書は、前回に引き続き今回も大きくアップデートされた。スコープは当然変わってはいないが、WHOのガイドラインの発表を見て、本文書も、最終稿に向けて体裁をWHOのフォーマットにしている。（表紙がITUではなくWHOのものになっている）

正直なところ、この文書が完成すれば他の文書は参考程度になってしまうかもしれない。

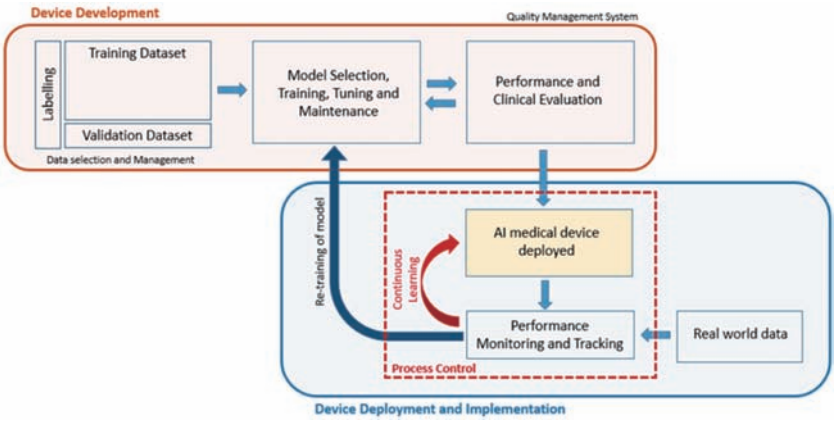
標準化の観点からもいくつか重要な項目が追加された。例えば、AIシステムの開発については、「AIシステムの設計、開発、トレーニング、検証、再トレーニング及び展開に関連するすべてのアクティビティは、ISO13485に基づいた品質管理システムの下で実行及び管理すること」という要求条件が明確にされた（**図1**）。

ISO13485は“Medical devices-Quality management systems-Requirements for regulatory purposes”という医療機器の品質管理に関するISO標準で、これによりWHOはAI4Hを医療システムとみなす、ということを明確にしている。これは、IMDRF“Software as a Medical Device（SaMD）：Key Definitions”に従って、AI4HをSaMDとみなしていることから当然と言える。しかしながら、個人の健康データのようなPHR（個人健康記録）をAIで処理するようなシステムが、SaMDとみなされるかどうかについては、議論が必要と思われる。何を医療機器とみなすかは、国によって異なっており、このWHOの要求条件がそのまま採用されるかどうかは不透明である。

また、AIワークフローのライフサイクルも米国のFDAの



■図2. AIワークフローにおけるAIシステムのプロダクト・ライフサイクル



■図1. AIシステムの開発と市場導入のプロセス

ものをそのまま採用している。

医療用サイバーセキュリティについても、IMDRFのものをそのまま引用している。

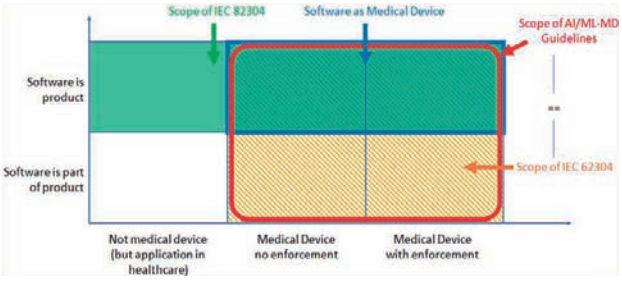
AI製品を購入する場合のガイドラインは英国のNHSのものを引用している。

このように既に存在するガイドラインを基に作成されているという点で「車輪を再発明しない（Not reinventing the wheel）」という標準化の理想から言って望ましいことではあるが、各国のガイドライン同士の整合性や、各国で異なる法律、規制との適用性など、本文書が公開されたのちにもさらに検討が必要と思われる。

2.2 DEL2.2：Good practices for health applications of machine learning：Considerations for manufacturers and regulators

本文書は、もともとは「AIベースの医療機器のガイドライン－規制要件」という名前だったのだが、いつのまにか変更され、メーカと監督官庁へのグッドプラクティス、ということになっている。内容は、AI特に機械学習（ML）を用いた医療機器（AI/ML-MD）開発の際の推奨プロセスに関する要件を記述し、AI4Hの実装開発者とメーカに指針を与えることを目的としている。この文書は、データ駆動型AIシステムのみを対象としており、非データ駆動型であるルールベースあるいはエキスパート・システム型のAIは対象としていない。

本文書は、(1) 包括的であること、及び/または、(2) 規制、指令、標準、または同様の法的拘束力のある規制の枠組みまたは地理的管轄区域のガイダンス文書となることを意図しない、とされている。しかし、監督官庁へのガイドラインにもなっているので、既にWHOが作成しているAI規制に関する文書との整合性が気になる。



■図3. 本文書のスコープ

このスコープ図（**図3**）から分かるように、DEL2.2のガイドラインは、特に医療用ソフトウェア（IMDRFで定義されているいわゆるSoftware as a Medical Device（SAMD））を対象としている。

ここで、IEC62304 “Medical device software−Software life cycle processes” は欧州連合と米国によって採用された医療品のソフトウェア設計に関する統一規格である。

また、IEC 82304, “Health software−Part 1：General requirements for product safety” は、「個人の健康管理・維持・向上目的または医療の提供に使用されることを意図したソフトウェアを対象とした標準で、いわゆる医療機器よりも広い範囲の機器を対象としている。

この文書は、大きく分けて、一般的な要求条件とメーカなどが市場導入する前後の要求条件に分かれている。一般的要件には、プロセス要件と能力要件がある。また、市場導入前の要件として、(1) 使用目的と利害関係者の要件、(2) 製品及びソフトウェア要件、(3) データ管理要件、(4) モデル開発要件、(5) 製品開発要件、(6) 製品検証要件、(7) 製品リリース要件、が挙げられている。また、市場導入後の要件には、(8) 生産、流通、設置要件、(9) 市販後調査要件、(10) 廃止措置要件が挙がっている。

また、Annex CにはRelationship to other guidelines and standardsとして、他の標準との関係を上げている。特に、IMDRFの「医療機器及びIVD医療機器の安全性と性能の基本原則」との関係を細かく書いていることは、参考になる。

2.3 DEL.3 AI4H要件仕様

本文書の目的は、一般的なAI4Hシステムの情報、機能、動作、運用についてのライフサイクルに基づいたシステム要件仕様（SyRS）を定義することとされている。具体的には、特に以下の2つを定義する：

(1) AI4Hシステムのシステム設計、システム検証、妥当性確認のプランニングと手順を提供するプラットフォームの定義。

(2) FG-AI4Hの他のWGとTGとの共同作業に則って、AI4Hシステム開発ライフサイクル中の要件追加、変更を識別、制御、追跡する機能を提供する。

このために、今回「要件トレーサビリティ・マトリックス」（単なるチェックリスト表）が提案された。そのテンプレート案を**図4**に示す。

このテンプレートは明らかに人間が記述することを想定しており、システム要求条件のチェックリストが人手で行われることは、AIシステムの要求条件として妥当であるか、

AI4H Topic Description to AI4H System Requirements Traceability Matrix-Template				
AI4H TDD Description	AI4H Requirements Traceability Attribute	TG-Symptoms (verification status)	TG-Ophthalmology (verification status)	TG-xxx
TDD Section 3.2 Benchmarking by AI developers				
Which scores and metrics have been used?	☐ Performance metrics			
	☐ Robustness metrics	✓		
	☐ Explainability metrics	✓		
	☐ Fairness metrics	✓		
	☐ Safety metrics	✓		
How did they approach the acquisition of test data?	☐ Protocols and tools to collect and integrate diverse data	✓		
	☐ Data acquisition / sensing modality			
	☐ Data acquisition / sensing device type			
	☐ Sensor device-OS version, firmware version			
	☐ Data collection place			

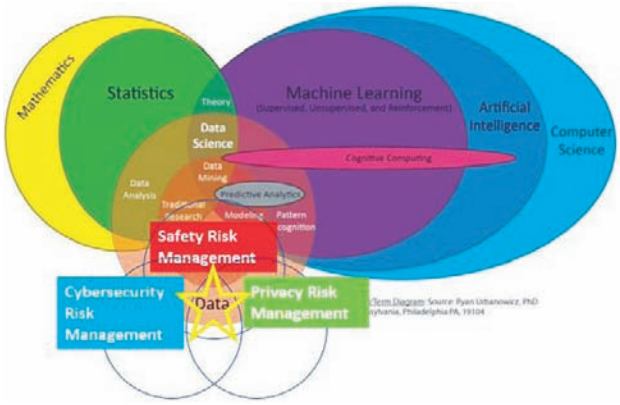
■図4. トレーサビリティ・マトリックス テンプレート案

再検討が必要と思われる。エディターによれば、本文書は80%以上完成しており、レビューに回すとのことである。当初の期待が大きかっただけに、残念な内容の文書と言える。

2.4 DEL 4 AIソフトウェアライフサイクル仕様

本文書は、AI4Hソフトウェアのライフサイクルに関連する標準とベストプラクティスの特定を目的としている。しかし今回は、サイバーセキュリティが特に加えられた。その理由は、ライフサイクルの監視は、リスク管理とサイバーセキュリティにとって重要であるということのようだ。

特に、健康医療においては、企業倫理協定（Corporate Integrity Agreement：CIA）によって、セキュリティとプライバシーが規定されることが関係している。それゆえ、臨床試験とヘルスケアAIアーキテクチャの設計段階でセキュリティとプライバシーの要件を含める必要があり、これはライフサイクルのスコープである。本文書では、NIST（米国標準技術研究院）のサイバーセキュリティに関するガイドラインを参照している。



■図5. SaMD データのライフサイクル

2.5 DEL 5：データ仕様

この成果物は、6つの個別の成果物のセットを包括的に組み合わせたもので、人工知能（AI）及び機械学習（ML）モデル／健康目的の方法に使用される場合のデータ仕様に関連する仕様を記述している。今回、アップデートが発表されたのは、5.3のデータアノテーション仕様と5.6のデータ共有だけで、残る5.1：データ要件、5.2：データ収集、5.4：トレーニングとテストデータの仕様、5.5：データ処理の各文書については、アップデートされなかった。

2.5.1 DEL 5.3 データアノテーション仕様

データアノテーションは、AI、機械学習では特に重要な分野である。データアノテーションの品質管理がAIの性能を決めると言っても過言ではない。前回の会合で、本文書はかなり完成度を上げ、今回はほとんど新しい追加がなかった。

今回、実装上の細かい規定も書かれており、例えば、出力ファイルには、アノテーション文書と元の画像が含まれる。アノテーション文書は、xml、json、txtなどの形式が想定されるが、それ以外でも可能である。アノテーション文書には、画像パス、オブジェクト名、オブジェクト座標の3つ以上のデータ項目が含まれていなければならない。また、照会の簡便さのために、画像とドキュメントの両方に同じルールに従ってファイル名を付けること、とされる。

医療で使われるEHRなどのテキストデータもAIにとっては重要な対象と言えるので、画像データだけを前提にしているように見えることが懸念点と言える。

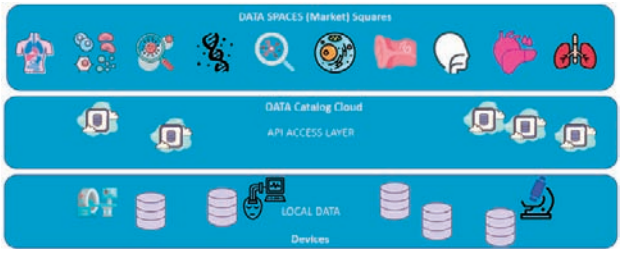
2.5.2 DEL 5.6：データ共有

本文書は、健康関連データを安全にデータ共有するための要件など、既存のベストプラクティスを概観することを目的にしている。また、分散環境とフェデレーション環境に基づいたソリューションが提案されている。

本文書はオープンコードイニシアチブの枠組みを使うことを提案している。

オープンコードイニシアチブは、AI専門家たちが使用できる規制に適合したプラットフォームを構築するためのツールとしてのソフトウェアを提供することを目的としており、そのためには、プラットフォームの開発者だけでなく、規制当局や医療専門家も要求条件という形で参加することになる。また、普遍的で国境を越えて使用できるツールを構築すること、そして複数ステークホルダーがアルゴリズムを構

築、評価、使用できることが可能である、とされる。まだ具体的な内容は詳しく書かれていないが、WebベースのAPIを提供することにより、分散的に存在するデータソース間でデータ共有が可能になるような仕組みについて、いくつかの方法を記述することになるだろう。図6にハイレベルのアーキテクチャを示す。



■図6. データ共有のアーキテクチャ

2.6 DEL 7. AI4H評価事項

本文書群は、FG-AI4HにおけるAI評価に関する出力文書をまとめたものである。

今回は、DEL.7.2、DEL.7.3、DEL.7.4がアップデートされた。

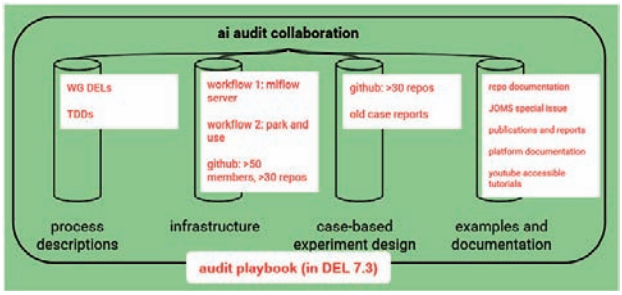
2.6.1 DEL 7.2 AI技術試験仕様

本文書では、AIをコンピュータを用いてテストする方法を指定することになっている。今回のアップデートでは、機能的試験と非機能的試験についての議論があった。

また、現在やるべきこととして、寄与文書の募集、成果物のAI技術指標仕様との調整、共同エディターの募集、https://www.nist.gov/standardsgovから標準についての情報を加える、となっており、現状では、まだ当初目指していた進捗ができていないと見える。

2.6.2 DEL 7.3 データと人工知能の評価方法

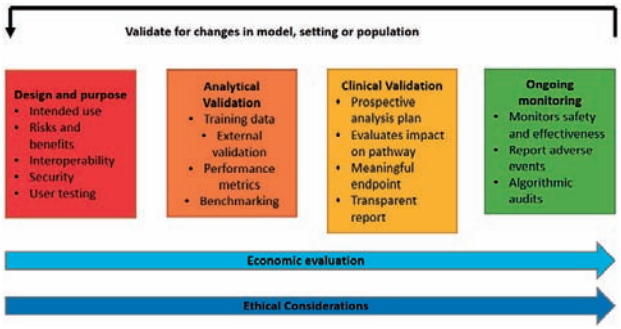
本文書は、もともとはAIベースの製品の様々な段階でアルゴリズムのバイアスを理解及び特定する方法の概要を述べることになっていた。しかし今回は、「アルゴリズム監査と品質管理」というタイトルと運用化評価（EP）及びレポーティング（RP）を特に取り扱う、という提案があった。その理由は、現在、AI分野が「信用度」という課題を抱えており、その原因がレポーティングにある、という評価が一般的になってきたことにある。FG-AI4Hには既にいくつかの試験仕様が存在する。これらを使ってAI監査の仕組みを作ろう、というのが、本文書の提案である。図7にそのプロセスの概要を示す。



■図7. 提案されたAI監査プロセス

2.6.3 DEL 7.4 AI4Hの臨床評価

本文書では、AI4Hシステムの臨床評価には、2つの分野横断的なテーマを持つ4つの主要なフェーズを含めることを推奨している（図8）。評価のフェーズとは、モデルの設計と目的の評価、分析的検証、臨床的検証、および継続的な監視である。分野横断的なテーマとは、倫理的な考慮、そして費用対効果評価である。



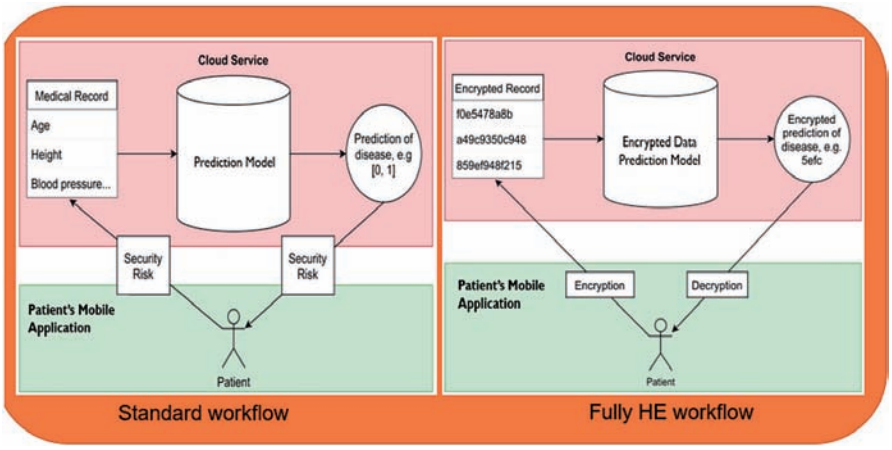
■図8. 健康医療におけるAI技術の評価のための枠組み

本文書では、これらのフェーズに分けて評価について議論している。

2.7 DEL 9.1 モバイルアプリケーション

本文書には、モバイルアプリケーションを使用したAI4Hツールの開発、試験、ベンチマーク等を記述することになっている。もともとはモバイルアプリだけで、クラウドを用いたアプリは別文書の予定だったが、今回、この文書のスコープにもクラウドを入れることが発表された。

今回のアップデートでは、セキュリティの課題についての提案があった。まず、プライバシー強化技術（PET）の必要性が指摘され、そのためのソリューションとして準同型暗号（Homomorphic Encryption：HE）を利用することが提案された。準同型写像（homomorphism）は構造を保持するので、HEは暗号文に対して操作を加えた場合、復号化された際に、平文に対して実行された操作を反映できる。



■図9. 準同型暗号でプライバシー強化されたAIモバイルアプリのワークフロー

したがって、個人データが、AIアルゴリズムの訓練や医療診断推論に使用される場合でも、安全であると主張されている。この仕組みを使ったモバイルアプリのワークフローを図9に示す。

今回、このような提案は成されたが、出力文書草案に反映されていないため、出力文書の進捗はなかった。

2.8 DEL 10 AI4Hの使用例：トピックの説明文書

第10のトピック文書は、現在24の分野が上げられている。

10.1 心血管疾患管理（TG-Cardio）
10.2 皮膚科（TG-Derma）
10.3 細菌感染と抗菌薬耐性の診断（TG-Bacteria）
10.4 高齢者の転倒（TG-Fall）
10.5 組織病理学（TG-Histo）
10.6 マラリア検出（TG-Malaria）
10.7 母子の健康（TG-MCH）
10.8 神経障害（TG-Neuro
10.9 眼科（TG-Ophthalmo）
10.10 アウトブレイク検出（TG-Outbreaks）
10.11 精神医学（TG-Psy）
10.12 AI for radiology（TG-Radiology）
10.13 ヘビとヘビの識別（TG-Snake）
10.14 症状の評価（TG-Symptom）
10.15 結核（TG-TB）
10.16 容積測定胸部CT（TG-DiagnosticCT）
10.17 歯科診断及びデジタル歯科（TG-Dental）
10.18 偽造医薬品（TG-FakeMed）
10.19 一次及び二次糖尿病予測（TG-Diabetes）
10.20 内視鏡検査用AI（TG-Endoscopy）
10.21 筋骨格医学用AI（TG-MSK）
10.22 生殖と妊娠のためのAI（TG-Fertility）
10.23 公衆衛生用AI（TG-Sanitation））
10.24 ポイントオブケア診断用AI（TG-POC）

3. 今後の予定と課題

FG-AI4H会合には、毎回50以上の寄与文書が寄せられ、今回の会合からの出力文書はリエゾン文書、レポートを含めて200近かった。関心の高さを物語っていると言える。しかしながら、文書数が多いため、細かい議論をする時間があまり取れているとは言えず、提案された文書がそのまま出力文書となっているのが現状であり、標準としてITUの勧告化のプロセスに乗せることができるか、若干不透明なところもある。

そのため、医療業界や医療機器メーカたちや他のAI団体とのコンセンサスを得られるかどうかが今後、課題となる可能性がある。

一方で、AI業界自体の進歩は速く、医療業界においてもAIがよりコモディティ化する可能性があるため、早急に標準化を進めることが必要と思われる。

またFG-AI4Hはフォーカスグループであるため、時限的な存在であり、いつまでも現在のような活動を続けることはできない。そのため、前回のSG16会合では、米国などの提案により、1年の限定延長が認められたが、その後は、FGとは別の枠組みで活動を続けることになる。

FG-AI4Hの次の会合は、2022年5月31日から6月2日に第15回会合（O会合）がドイツのベルリンで開催されることになっている。コロナ禍が始まってからの最初の対面による会議となる。