



ITU-T FG-AI4H会合報告

慶応義塾大学大学院 政策メディア研究科 特任教授

かわもり まさひと
川森 雅仁



1. はじめに

ITUがWHOと協力して開催している Focus Group on AI for Health (FG-AI4H) は、2018年7月9日～20日にスロベニアのリュブリャナで開催されたITU-T SG16において設立された。最初の会合をジュネーブのWHOで2018年10月25日～27日まで開催してから、今回で10回目のFG会合になる。今回は前回に続きコロナ禍のためオンライン会議となり、9月30日～10月2日まで開催された。オンラインであることから、ヨーロッパや米国だけでなく、中国、韓国をはじめ、アフリカやインドからの参加者も増えている。

2. 今次会合での合意事項・結果

これまでに、FG-AI4Hの成果物のリストが計画され、それに対応するグループが設立された。既にその大体の概要が明らかになってきた。FGの出力文書は2種類に分かれる。第1種は、より一般的な共通事項をまとめた成果物で、9つの領域での一般的な要求条件を中心に記述されている。

それに対して、第2種の出力文書は、20余りの特定健康医療領域のトピックに関するグループ (TGと呼ぶ) の出力文書である。これらは、お互いに重複を避け、専門性を維持しながら、様々なユースケースを収集し、それに対応するAIや機械学習 (ML) のベンチマークのためのタスクを記述するとされている。

これら2種類の文書を合わせると、少なく見積もっても30、分冊も含めると50ほどの出力文書になることが予想される。以下にこれらの概略を述べる。

3. 共通項目出力文書案

これらの文書群は、AI4H全般に関わる、より一般的な共通事項に関する成果物で、倫理、規制、要件、ソフトウェア・ライフサイクル、データ仕様、モデルトレーニング、モデル評価、拡張、アプリ、の9つの領域での一般的な要求条件を中心に記述されている。

以下、これらの出力文書を概述する。

D1 AI4Hの倫理に関する考慮事項

デジタル技術、機械学習、AIは、医学、研究、公衆衛

生の分野に前例のない方法で革命をもたらしている。この急速に発展している分野は、大きな期待を抱かせると同時に、倫理的、法的、社会的懸念を受けている。AI技術の倫理的開発と使用に関する多くの課題は未解決のままである。グローバルヘルスにおけるAIの設計と実装のための調和のとれた倫理ガイダンスを策定することは緊急の必要性があり、また、世界規模でAIのメリットを確保するには、新しい共同研究体制の確立が必要である。このような背景から、この文書は、AI4Hの使用に関する主要な規制上の考慮すべき事項の概要と、AI4Hソリューションの使用に関する開発者と規制当局間が共通理解を確立するために議論すべき対象を記述している。しかしながら、この文書自体はガイダンス、規制、またはポリシーとなることを意図したのではなく、規制当局、開発者、あるいはその他の利害関係者が医療健康分野でAIを利用しようとする時に考慮すべき倫理的事項についての情報源となることを意図している。

D2 AI4H規制のベストプラクティス

AIソリューションの安全な適用に向けたAI4Hの規制に関するベストプラクティスに関するガイドライン。これには、様々な業種にわたるAI規制の共通要素と、各国の固有の規制が含まれる。

D2-1 健康ソフトウェアのためのAIへのIMDRFの基本原則のマッピング

国際医療機器規制当局フォーラム (IMDRF) は医療ソフトウェア (Software as a Medical Device: SaMD) の規制枠組みとして“Essential Principles of Safety and Performance of Medical Devices and IVD Medical Devices”, IMDRF Good Regulatory Review Practices Group, IMDRF GRRP WG/N47 FINALを 2018年10月31日に発表している。

しかしながら、AI4HソフトウェアはこのIMDRF Essential Principles (EPs) が作成された時点では、SaMDとして想定していなかった新しい機能を多く提供することになる。そのため、この出力文書は、AI4HソフトウェアをIMDRF EPにどう対応 (マッピング) させるかについて推奨することになる。SaMDとして規制されるべきすべての面が、AI4Hソフ

トウェアに適用されるべきかどうか、また、もし適用できるならば、どのように適用できるかという問題を扱っている。

D2-2 AIベースの医療機器のガイドライン—規制要件

この文書は、規制当局及び医療機器メーカーに対して医療機器向けAI (AI-MD) が順守すべき包括的要件を分析可能にするベストプラクティスを示し、その工程についての共通理解を持つための一連のガイドラインを定義する。

ここで、AI-MDの規制範囲には、以下が含まれる。

(a) 規制対象及び非規制対象の医療機器

(b) 規制の有無に関わらない医療機器

ただし、この文書は、AI及び機械学習製品の技術的側面のみを含み、市場戦略的位置付け、市場評価、収益性などの商業的側面は含まない。

AI-MDの製品に含まれるのは以下である。

(c) 医療機器としてのソフトウェア (SaMD)

(d) 医療機器内のソフトウェア (Software in a medical device : SiMD)

(e) 医療行為または医療システムの効率を改善することを目的とした医療アプリケーション

また、以下はAI-MDの製品範囲に含まれない。

(a) 医療施設の管理サポート

(b) 健康増進・維持のためのソフトウェアアプリケーション

この文書に含まれる一連のガイドラインは、以下を意図していない。

1) 包括的であること

2) いかなる地理的管轄区域においても、その地域の規制、指令、標準、あるいは同様の法的拘束力のある規制の枠組みまたはガイダンス文書等の代わりとなること

D3 AI4H要件仕様

この文書は、一般的なAI4Hシステムの情報、機能、動作及び運用等を記述するシステム要件仕様(SyRS)を定義する。

- SyRSのスコープには、検討中のAI4Hシステムの情報、機能、動作及び運用面を定義する要件モデルが含まれている。具体的には次のとおり。

—AIソフトウェア要件及びAIの実行すべきタスクを曖昧なく定義する方法に関するベストプラクティス。これには、使用目的の明確な説明が含まれる。

—AI4Hソフトウェアを既存の介入との関係でどう分類するかに関する手順。(AI4Hソフトウェアが、既存の健康介入ワークフローを置き換えることになるか等)

—リスク管理ガイドライン

- このSyRSは一般に適用可能であるべきで、FG-AI4Hの特定の専門分野を扱っているトピックグループのすべてに適用可能である。トピックグループの特定の要件に対応するために変更、カスタマイズ、拡張などを可能にするべき。
- 要件仕様は、ユースケース、グラフィカルな方法、数学モデル、文章化など、またはこれらの組合せから定義。

D4 AI4Hソフトウェア・ライフサイクル仕様

この文書は以下をスコープとする。

- AI4Hソフトウェアのライフサイクルに関連するすべての標準とベストプラクティスを特定すること。(他のソフトウェアのライフサイクルと同様に、AIソフトウェアのライフサイクルを指定する必要がある。)
- ギャップ・アナリシス(既存の文書の要約と批判的レビュー)。
- AI4Hヘルスソフトウェアに特徴的なライフサイクルステップの特定(例; アノテーションが必要な可能性のあるデータに基づくトレーニングやテスト手順)。
- AI4Hヘルスソフトウェア・ライフサイクルの仕様と、1つの文書内の様々なライフサイクルステップのベストプラクティスの定義。ベストプラクティスの概要と例。

D5 データ仕様

この文書は、6つの文書を組み合わせたもので、AI及び機械学習(ML)モデルが健康目的に使用される場合のデータ仕様に関連する6つの重要項目に対応する。

D5.1 データ要件

この文書には、FG-AI4Hに送付されたデータの受け入れ基準がまとめられており、その基準に準拠するための原則と規則が記載されている。AI4Hのベンチマーク枠組みの核となるのは、AI開発者がアクセスできない、各特定医療領域(後に述べるTGの領域)のユースケースごとのテスト用非公開データセットであるので、これらの原則は非常に重要。

D5.2 データ収集

この文書には、どのような国や国際的医療機関でも簡単に採用できるように、標準プロトコルに基づく公的医療データの取得及び管理モデルのフレームワークを記述している。この文書は、医療施設での電子カルテ(EHR)等の基本的デジタル化を前提としている。公衆衛生のための標準化されたEHRの使用には、患者の個人情報を匿名化することによるプライバシー保護の問題、同じ情報の



細かい変更を含む複数レコードの融合等の課題がある。

D5.3 データ・アノテーション仕様

データ・アノテーションは、モデルのパフォーマンスの最も信頼できる要素の1つであり、AI4Hのデータ品質管理にとって重要な機能を担う。本文書は、定義、背景と目標、フレームワーク、標準操作手順、シナリオ分類と対応する基準及び推奨されるメタデータなどを含む、データ・アノテーション仕様の一般的なガイドラインを提供することになっている。

D5.4 トレーニングとテストデータ仕様

この文書は、機械学習モデルのトレーニングとテストで使用されるデータセットの技術要件仕様を体系的に記述する。機械学習プロセスのライフサイクルのトレーニング及びテストデータ準備フェーズでのデータ誤りの可能性を最小限に抑えることを目的としたデータ品質保証のベストプラクティスについて説明する。

D5.5 データ処理

この文書では、データが受け入れられた後のデータ処理方法について概説する。健康データは、重要で機密性の高いデータである。この種のデータの処理は、多くの場合、データ保護法によって定義された厳密なフレームワークに関連付けられている。データの提供者だけでなく、すべての利害関係者の間でFG-AI4Hの信頼を確保するために、厳格なデータポリシーを設定することが重要である。データ処理ポリシーは特に次の問題に対処する必要がある。

- (a) 個人の健康データの使用規制の遵守
- (b) モデル評価の目的でFG-AI4Hが保持している非開示テストデータを非開示とすること

D5.6 データ共有の実践

この文書は、健康関連データのデータ共有に関する現存するベストプラクティスの概要を記述し、安全なデータ共有を可能にするために必要なすべての手順と要件を説明している。この文書では、データプロバイダー、データプロセッサ及びデータレシーバーという3つの役割を規定している。また、既に確立されたデータ共有方法に加えて、分散環境とフェデレーションに基づいた新しいプライバシー保護AI/MLモデルの概要を記述している。

D6 AIトレーニングのベストプラクティス仕様

この文書は次の2つの部分で構成されている。

- (1) AIモデルのトレーニング中に遭遇した課題の要約と、いくつかの推奨される方法を提案。

- (2) トレーニングされたAIモデルの透視的なレポートに使用できる推奨フレームワーク。

D7 AI4H評価事項

この文書では、新規または不可欠なAI4H評価特性を特定し、AI4Hモデル評価のベストプラクティスを選択した文献等から収集し、ベンチマークプラットフォームの要件を検討する。これにより、AI4H評価プロセス記述、技術試験仕様、測定基準、臨床評価などが、さらに細かく記述される。

D7.1 AI4H評価プロセス記述

AI4H評価プロセス記述は、最先端のAI評価のための原則と方法の概要を記述すると同時にAI4Hの評価プロセスの嚆矢としての役割がある。このプロセスの説明には、現状の評価原則と評価法、また、AI4Hに特有の評価に関する要件、ソリューションの見直しや、AI4H評価が持つ倫理上のリスク面についても検討する。さらに、AIの理論的基盤に基づいて、現在の評価方法がREAL AI概念に向かってどのように進化するかについての洞察を得ることができる。

D7.2 AI技術試験仕様

この文書では、コンピュータによる (in silico) AIの試験法とその要求条件を規定している。その他には、AIチャレンジ等から得られた試験手順のベストプラクティスも参考として挙げられる。さらに、AIアプリケーションに特化していない重要なテストパラダイムについても言及する必要がある。

D7.3 データと人工知能の評価方法 (DAISAM) リファレンス

ここでは、AIを用いた製品が実際の臨床環境に適用された際に重大な影響を与える可能性があるアルゴリズム上のバイアスを理解及び特定する方法の概要を示す。その目的は、各グループの最も正確なモデルをトレーニングすることを患者の少数グループに害を与えることなく実行可能にすることにある。また、眼前の問題に応じてバイアスを軽減する方法も示す。

D7.4 AI4H臨床評価

FG-AI4Hの目的の1つは、規制当局や意思決定者に対して、AI4Hの「最新バージョン」の標準が実運用状況やユースケースに適していることが明確になるような中間ステップを、速やかに信頼できる形で提供することである。これには、独立して収集された、高品質で安全な（つまり、これまでにシステムに提供されたことがない）総合的かつ代表的なデータセットを作成する必要がある。



D8 AI4Hの拡張と採用

この文書はまだ内容が確定していない。

D9 AI4Hアプリケーションとプラットフォーム

この文書は、モバイルアプリケーションとクラウドベースのAIアプリケーションを使用したAI4Hツールの開発、それらのテストとベンチマークに関する一連の規則案である。

D9.1 モバイルアプリケーション

この文書は、モバイルアプリケーションを使用したAI4Hツールの開発、テスト、ベンチマークに関する一連の規則を記述する。

- モバイルアプリケーションを使用してAI4Hツールを開発するためのルールを準備
- AI4Hを使用したモバイルアプリの規制/倫理ルールについての議論
- 様々な利害関係者間のオープンなコミュニケーションのためのフォーラムを提供
- アプリのベンチマークを開発
- FG及びWGと協力して、プロセス全体を調整

D9.2 クラウドベースのAIアプリケーション

この文書には、クラウドベースのAIアプリケーションの開発、それらの試験及びベンチマークに関する一連の規則案が含まれている。

- クラウドベースのAIツールに関連する技術、セキュリティ、法的な問題等についての議論
- 様々な利害関係者間のオープンなコミュニケーションのためのフォーラムの提供
- FG及びWGと協力してベンチマークプロセスを調整

4. 特定分野トピック文書案

上記のAI4Hの共通項目に関する出力文書に対して、特定の健康医療領域に関して20余りのトピックグループが抽出され、お互いに重複を避け、専門性を維持しながら、様々なユースケースを収集し、それに対応するAI/MLベンチマークタスクを記述する。以下にそのトピック分野を示す。

- 1 心血管疾患管理 (TG-Cardio)
- 2 皮膚科 (TG-Derma)
- 3 細菌感染と抗菌薬耐性の診断 (TG-Bacteria)
- 4 高齢者の転倒 (TG-Fall)
- 5 組織病理学 (TG-Histo)
- 6 マラリア検出 (TG-Malaria)
- 7 母子健康 (TG-MCH)

- 8 神経障害 (TG-Neuro)
- 9 眼科 (TG-Ophthalmology)
- 10 アウトブレイク検出 (TG-Outbreak)
- 11 精神医学 (TG-Psy)
- 12 放射線科 (TG-Radiology)
- 13 ヘビとヘビの識別 (TG-snake)
- 14 症状の評価 (TG-Patho)
- 15 結核 (TG-TB)
- 16 CT (TG-DiagnosticCT)
- 17 歯科診断及びデジタル歯科 (TG-Dental)
- 18 偽造医薬品 (TG-FakeMed)
- 19 糖尿病 (TG-糖尿病)
- 20 内視鏡検査用AI (TG-内視鏡検査)
- 21 筋骨格医学のためのAI (TG-MSK)

それぞれの分野について、以下の情報が提供されることになっている。

- タスク (分類/検出/セグメンテーション/予測など)
- 参照規範 (最先端のタスク介入方法)
- 入力データタイプ (テキスト/画像/ビデオ/オーディオ/数値/その他)
- テスト/トレーニング用データセット (公開データセット/自分で収集など)
- データ・アノテーション (手順/注釈者番号/ツールなど)
- アルゴリズム (このTGで使用される特定のモデル)
- 評価 (このTGで使用されるメトリック)

5. 今後の予定と課題

FG-AI4Hの次の会合は、2021年1月27日～29日まで、コロナ禍のため引き続きオンラインでのバーチャル会議となる。ワークショップの併催も予定されている。

今回の会合で、出力文書の概要が明らかになってきたので、その議論が中心になると思われる。特に20余りの特定領域の文書化が進んでいないので、その議論が進むことが期待される。また、WHOはAI4Hの倫理面での規定の議論が進むことを望んでおり、それに関する議論も進められるということが予想されている。

一方で、上で見て分かるように、文書の数がかなり多くなっているだけでなく、多岐にわたっていると同時に、いくつかの重複がみられる (例えば、テストや評価法についての部分)。また、モバイル・アプリやクラウド・アプリについては、実装に関わることであるため、どういうことを標準化するべきかについても議論が必要であろう。