

ヘルスケアSWG活動報告

～ヘルスケア領域におけるTrustworthy AI（信頼できるAI）の実現を目指して～

ヘルスケアSWG

Ubie株式会社 アクセラレーター本部代表・政策渉外参事/ JaDHA WG4リーダー
井上 真夢

Ubie株式会社 Chief AI Officer (CAIO)
風間 正弘

2026年10月7日

AISI事業実証WG 2026年度上期報告会

AISI Japan
AI Safety Institute

ヘルスケアSWGのリーダーを務めるUbie株式会社から2名で登壇します。



井上 真夢
Inoue Mamu

Ubie株式会社 アクセラレーター本部代表/政策渉外参事
日本デジタルヘルス・アライアンス WG4 SuBWG-B リーダー
日本医療ベンチャー協会主幹

- 2014年 総務省入省。電気通信事業分野の消費者保護や郵政行政、地方の情報通信施策振興やデジタル田園都市国家構想推進などの政策に8年間携わる。
- 2022年 ヘルステックスタートアップのUbie株式会社に入社。ビジネスパートナーアライアンスなど事業開発チームを経て、Public Affairs（政策渉外）担当に。日本デジタルヘルス・アライアンスでは生成AI活用ガイドの策定をリーダーとして牽引。

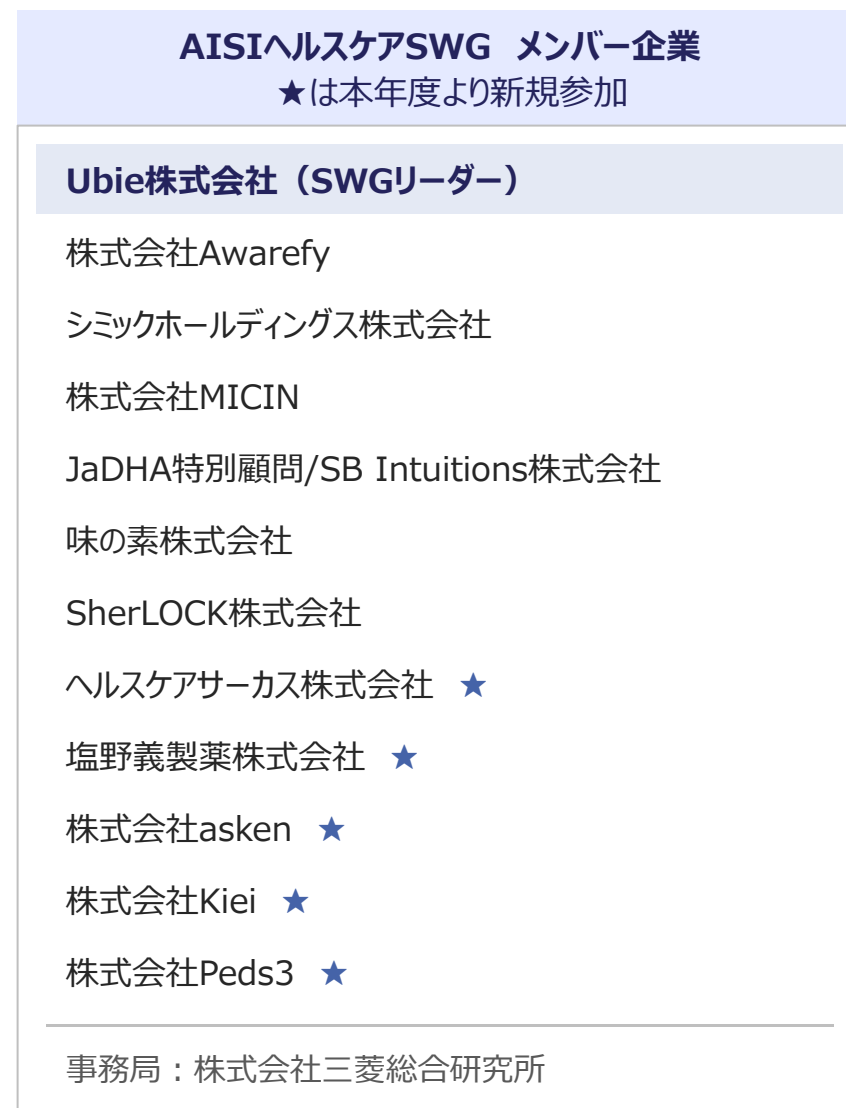


風間 正弘
Kazama Masahiro

Ubie株式会社 Chief AI Officer (CAIO)
津田塾大学 非常勤講師
国立国語研究所 外来研究員

- 2015年 東京大学大学院を卒業、リクルートホールディングス入社。様々な領域のデータ分析や機械学習アルゴリズム開発を担当。2018年よりIndeedに異動。
- 2020年 ヘルステックスタートアップのUbie株式会社に入社。AI問診の開発チームをリードし、2023年から生成AI活用を担当。
- 2018年 Forbes 30 Under 30 Japanを受賞
- 2022年 推薦システム実践入門(オライリー・ジャパン)執筆
- 2022,23年 東京都立大学非常勤講師

ヘルスケアSWGの体制・参加組織



日本デジタルヘルス・アライアンス（JaDHA）の概要



組織概要

会長：三友周太（シミックホールディングス株式会社）
製薬デジタルヘルス研究会および日本DTx推進研究会を統合し、
2022年3月14日に設立。



会員企業

約100社が参加
大手医薬品・医療機器メーカー、デジタルヘルスベンチャー企業、大手ICT企業等



設立背景

コロナ禍は社会におけるデジタル化の重要性が一層認識される契機に。先進的プログラム医療機器の実用化を促す施策の検討が進む。
「デジタルだからこそその価値」の評価、柔軟性のある制度・規制の実装が重要。



活動内容

業界の垣根を超えた横断的研究組織の組成と活動により、産業の発展、関連サービスや技術の普及促進を阻害する課題を深く洞察、デジタルヘルス産業の発展を巡る課題解決の在り方を提言する。

活動概要

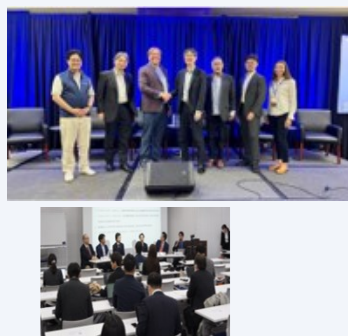
ビジョン・基本方針の策定



業界自主基準の策定



国内外の産学官連携



委員会体制 ※カッコ内は委員長企業

薬事・診療報酬制度の在り方検討委員会	テルモ
SaMD QMS検討委員会	塩野義製薬
SaMDにおけるデータ利用の在り方検討委員会	DeNA
生活者や患者に向けた広告規制の在り方検討委員会	事務局
デジタル医療サービスの円滑な利活用にむけた基盤構築委員会	NTT data
デジタルヘルスサービスの認知向上委員会	シミックホールディングス
信頼できるAIのイノベーション推進委員会	Ubic

- ヘルスケア領域は、生命・身体への影響が及ぶリスクやプライバシー性の高い情報を多く取り扱う分野であることから、これらを踏まえたAIセーフティ評価の在り方の検討が必要。
- ヘルスケア領域におけるTrustworthy AI（信頼できるAI）の社会実装に向けてガイド策定を開始。

生成AIを活用したプロダクトの広がり



生成AI技術の急速な発展により、ヘルスケア領域での活用が広がっている。



BtoB 医療従事者向けの文書作成支援・業務効率化ツール



BtoC 健康相談チャットボット・メンタルヘルスケアアプリ

医師の働き方改革や患者コミュニケーションの向上が期待され、社会的・経済的価値をもたらすプロダクトが実用化。

ヘルスケア領域における現状



生命・身体・精神等に直接影響し得るヘルスケア領域では、ハルシネーション・説明可能性の欠如・プライバシー侵害などのリスクが顕在。



国内外でルールメイキングは進むものの、「何をもって信頼に足るか」の実践的な評価基準はまだ整っていない。

【ヘルスケア領域におけるAIセーフティ評価観点ガイドの策定】

日本のAI事業者ガイドライン等の指針やAISII策定の「AIセーフティ評価観点ガイド」や2025年12月に閣議決定されたAI基本計画が掲げる「Trustworthy AI」の実現に向け、AISIIの10項目の評価観点をベースにヘルスケアに特化したAIセーフティ評価の在り方を検討、開発フェーズ別のチェックポイントを提供し、イノベーションと安全性の両立を実現することが目的。

- Trustworthy AIの実現に向けた国内外の潮流を踏まえ、ヘルスケア領域に特化したAIセーフティ評価の実践的な指針を提供。マークダウン形式でも公開。
- 実用性を追求した設計とし、AISIIの評価観点に加えて実践を想定したフェーズごとの評価項目を整理。

ガイドの主な構成

1. 本ガイドの背景と目的

- ◆ ヘルスケア領域におけるAIセーフティの重要性
- ◆ ガイドのスコープ（AI提供者、Non-SaMD、テキスト生成）

2. ヘルスケア分野におけるAI動向

- ◆ 技術動向、市場動向、政策・業界動向

3. ヘルスケア領域におけるAIセーフティ評価の10観点

- ◆ 各観点の概要、想定リスク、実際の事例、評価項目例

4. AIプロダクト開発におけるAIセーフティ評価の実践

- ◆ フェーズごとの主な評価観点と具体的な実践方法
- ◆ 確認すべき事項のチェックリスト



- 2025年度は、AI提供者かつNon-SaMD（医療機器プログラムに該当しないAIプロダクト）、テキスト生成AI（LLM）を対象に検討。ユースケースはBtoB（医療機関向け）/BtoC（生活者向け）を想定。
- サービスの企画・開発・運用・リスク評価に携わる経営層やプロダクトマネージャー、エンジニア・法務等を想定読者として想定。

本年度のスコープ

対象者

AI提供者（学習済みの生成AIモデルをAPI経由等で利用してプロダクトやサービスの開発を行う事業者）

対象のプロダクト

非医療機器プログラム（Non-SaMD）

対象の生成AI

テキスト生成AI（LLM）（画像生成AI・音声生成AI等は本ガイドの対象外とする）

対象のユースケース

カテゴリ	ユースケース例
BtoB （医療機関向け）	文書作成支援、情報検索・要約、カルテ入力補助、患者説明資料の作成支援、医療文献の検索・要約 等
BtoC （生活者向け）	健康相談チャットボット、セルフケア支援、メンタルヘルスケアアプリ、服薬リマインダー、健康管理アプリ 等
その他	製薬企業向け情報提供支援、介護記録作成支援、臨床研究の文献レビュー支援 等

- ヘルスケア分野におけるAIの最新動向を医療特化型LLMをはじめとする活用技術の現状と活用領域を概観したうえで、国内外のAI活用プロダクトの市場動向と具体事例を紹介。
- 各国の政策・業界団体・国際機関によるAIセーフティへの取組を整理し、ヘルスケア分野に特化したAIセーフティ評価の具体的な事例についても取り上げている。

国内外のAI技術動向

- ・ 専門家の63%がAIを積極導入（NVIDIA調査）
- ・ **英語系医療LLM**
 - MedLM（Google）、Med-Gemini、Meditron-70B（EPFL）等
- ・ **日本語系医療LLM：SIP第3期を軸にアカデミアが主導**
 - 東大 相澤研：SIP-jmed-llm・JMedBench
 - 東大 松尾・岩澤研 × ELYZA：ELYZA-LLM-Med
 - 奈良先端大 荒牧研：大規模医療用辞書「JMED-DICT」
 - NII：モデル開発・NTCIRやAnswerCarefullyデータセット整備

国内外のAI政策・業界動向

各国動向

- ・ 日本：AI推進法（2025）・AI事業者ガイドライン・AISII設立
- ・ 欧州：AI Act（2024）リスクベース規制・汎用AI行動規範
- ・ 米国：America's AI Action Plan
- ・ 英国：AI法案・AI Security Instituteへ改名・方針転換

業界動向

- ・ 日本：JaDHA・HAIPによる生成AI活用ガイドライン策定
- ・ 米国：米国医師会レポート・CHAI「責任あるAIガイド」
- ・ 英国：MHRAによる規制サンドボックス「AI Airlock」実施

第3章：ヘルスケア領域におけるAIセーフティ評価観点について

- AISI策定のAIセーフティ評価の10観点について、ヘルスケア領域に適用した場合を以下の点から整理。
 - 想定されるリスクの例：各観点ごとにヘルスケア領域で特に懸念されるリスク
 - 実際の事例：国内外で報告されている関連事例を紹介
 - 評価項目例：サービス・プロダクトの企画・開発・運用において確認すべき評価項目を例示

No .	評価観点	ヘルスケア領域におけるリスク概要
1	有害情報の出力制御	医療・健康に関する危険な情報（自傷・暴力の助長、根拠を欠く治療法等）が出力され、患者の生命・健康に直接的な被害をもたらすリスク
2	偽誤情報の防止	ハルシネーションにより架空エビデンスや誤った薬剤情報が生成され医療現場や生活者に影響を及ぼすリスク
3	公平性と包摂性	特定の属性の患者に対しAIの精度や品質が低下し、医療における既存の格差が拡大するリスク
4	ハイリスク利用対処	非SaMDが事実上の医療機器として利用される「目的外利用」により法規制違反等が生じるリスク
5	プライバシー保護	要配慮個人情報を含む医療データが漏えい・不正利用され、患者のプライバシーが侵害されるリスク
6	セキュリティ確保	プロンプトインジェクション等の攻撃により、医療情報の改ざんや機密データの漏えいが生じるリスク
7	説明可能性	AI出力の根拠が不透明なまま臨床判断に用いられ、誤った医療行為や患者の不信につながるリスク
8	ロバスト性	方言・略語・非標準的な医療用語等の多様な入力に対し出力品質が不安定となり、誤った判断を招くリスク
9	データ品質	不正確または陳腐化した医療データに基づく出力が、臨床現場で誤った判断の根拠として用いられるリスク
10	検証可能性	事後検証や第三者監査が困難な状態では、問題発生時の原因究明ができず、社会的信頼を損なうリスク

想定リスク：

医療・健康に関する危険な情報（自傷・暴力の助長、根拠を欠く治療法等）が出力され、患者の生命・健康に直接的な被害をもたらすリスク

想定され得るリスクの例

- 人の生命・身体に直接的な危害を及ぼす行為を容易化・助長する情報
- 医学的に危険な情報（健康被害・医療機会逸失のリスク）
- 心理的依存と社会的孤立
- 基本的人権・尊厳の毀損

実際の事例

- **摂食障害支援チャットボット（Tessa）の停止：**
2023年、摂食障害患者向けに導入されたAI（NEDA支援団体のTessa）が、ユーザーに対し症状を悪化させる恐れのある「具体的なカロリー制限」や「減量方法」を推奨し、運用停止に追い込まれた事例※

フェーズ	評価項目例
①プロダクト設計	有害情報のリスク類型と許容基準が定義されていること
②モデル選定	基盤モデルの選定において有害出力の抑制能力が評価されていること
③プロダクト実装	入力から出力に至る多層的な有害情報の防止機構が実装されていること
④プロダクト検証	有害出力の抑制が実証的に検証されていること
⑤プロダクト導入運用	本番環境において有害出力が継続的に監視され、迅速に是正されること

- 実務に即した形でプロダクト開発を5つのフェーズに分類し、各フェーズにおいて重要となる評価項目を整理
- 各フェーズで重要となる要素を実務で活用できる粒度で解説
- チェックリストの形式に落とし込むことで、各事業者によるガイド活用の促進を企図

フェーズ	概要	主な評価観点
①プロダクト設計	・ プロダクトの目的・ユースケースの明確化、リスク評価、ガバナンス体制の構築	・ リスクアセスメント、法規制遵守、プライバシー・セキュリティ
②モデル選定	・ 用途に適したモデルの選定と安全性評価	・ モデルの安全性・性能評価、ベンダー信頼性
③プロダクト実装	・ システムアーキテクチャ、プロンプト設計、ガードレール実装	・ 入出力制御、ハルシネーション対策、透明性確保
④プロダクト検証	・ 総合的なテスト・検証とリスク評価	・ レッドチームテスト、バイアス評価、ドメインエキスパートによる検証
⑤プロダクト導入運用	・ 本番環境でのモニタリングと継続的改善	・ モデル変化管理、インシデント対応、利用者フィードバック

第4章：AIプロダクト開発におけるAIセーフティ評価の実践について

- 10評価観点と5フェーズをかけ合わせたマトリクスで整理
- 各評価観点軸と各フェーズ軸ごとに、リスクや対応策を整理し、実務にて参照できるように整理
- 幅広く活用できるように「プロンプトとスキルによるAIセーフティ評価の実践例」を公開

評価観点	① プロダクト設計	② モデル選定	③ プロダクト実装	④ プロダクト検証	⑤ プロダクト導入運用
有害情報の出力制御	有害情報のリスク類型を定義し、対応方針を設計	安全性ベンチマーク等で有害出力の抑制能力を評価	入力・モデル・出力の多層防御を実装	レッドチーミング・専門家レビューで抑制効果を検証	有害出力の発生状況を継続監視し、迅速に是正
偽誤情報の出力・誘導の防止	ハルシネーション等のリスクを類型化し、許容基準を定義	事実整合性・医療特化ベンチマークで正確性を評価	RAGによる根拠付けと出典明示の仕組みを実装	ハルシネーション率・出典正確性を定量的に検証	ハルシネーション率を継続監視し、参照データを最新化
セキュリティ確保	LLM固有の脅威を含むセキュリティ要件を定義	プロバイダーの体制とモデルの攻撃耐性を評価	インジェクション防御・認証・暗号化等を多層実装	ペネトレーションテスト・レッドチーミングで検証	脆弱性情報の収集・異常監視を継続的に運用

第4章：AIプロダクト開発におけるAIセーフティ評価の実践について

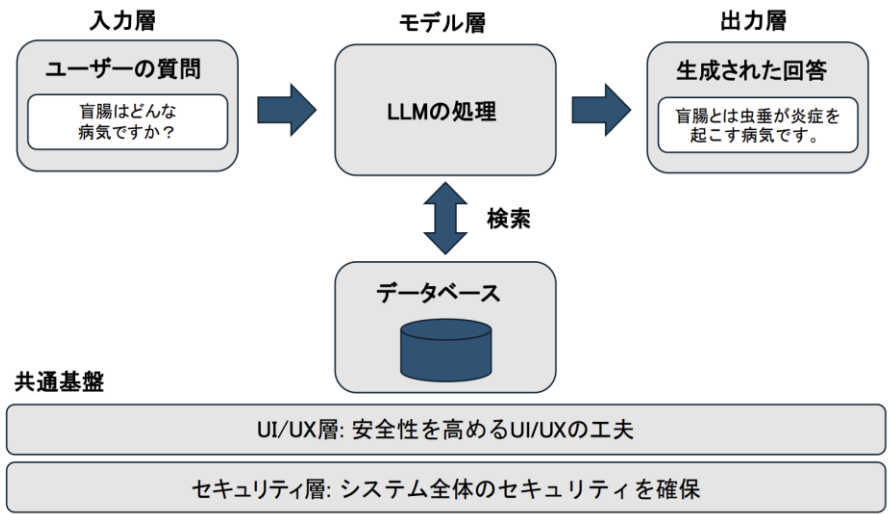
チェックリストの例、一部抜粋（フェーズ① プロダクト設計 セーフティ・バイ・デザインを重視した内容）

カテゴリ	確認事項	関連する評価観点	対応状況
全体設計	プロダクトの目的・対象ユーザー・ユースケースを明確に文書化しているか	横断	☐
全体設計	AIの役割の範囲と限界（やってはいけないこと）を定義しているか	ハイリスク利用	☐
リスク定義	有害情報のリスク類型（不正確な医療情報、自傷誘発、心理的依存の助長等）と許容基準を定義しているか	有害情報	☐
リスク定義	ハルシネーション・偽誤情報のリスク類型と許容基準を定義しているか（特に投与量・禁忌等の致命的領域）	偽誤情報	☐
リスク定義	想定される入力の多様性（表記ゆれ、方言、OCR誤認識等）を洗い出し、出力一貫性の許容基準を定義しているか	ロバスト性	☐
設計原則	プライバシー・バイ・デザインの原則を適用しているか	プライバシー	☐
設計原則	セキュリティ・バイ・デザインの原則を適用しているか	セキュリティ	☐
ガバナンス	医療・セキュリティ・法務を含む多職種チームを編成しているか	横断	☐
ガバナンス	セーフティレビューを経ずにリリースが行われない仕組みを設計しているか	横断	☐
法規制	適用される法令・ガイドライン（薬機法、医師法、個人情報保護法、3省2ガイドライン等）を特定し、対応要件を整理しているか	横断	☐
検証計画	事後検証に必要なログ要件と評価体制の計画を策定しているか	検証可能性	☐

例) フェーズ3 プロダクト実装のセーフティ対策内容

- モデル単体だけでなく、プロダクト全体として多層的に対応し、顧客に安心・安全に価値を提供できるプロダクトを開発

ヘルスケア領域におけるAIプロダクトの一般的なアーキテクチャ
(例)



レイヤー	概要	安全性対策のポイント
入力層	ユーザーからの入力を受け付け、モデルに渡す前の処理を行う	入力フィルタリング、プロンプトインジェクション対策、個人情報マスキング
モデル層	LLMによる推論処理を行う	システムプロンプト設計、パラメータ設定、構造化出力
出力層	モデルの出力を加工し、ユーザーに提示する	出力フィルタリング、ガードレール、引用元明示
データベース (RAG)	外部データを検索し、モデルの応答精度を向上させる	検索品質の向上、データ品質管理、アクセス権限制御
UI/UX層	ユーザーとのインターフェースを提供する	安全性を高めるUI設計、免責、利用規約
セキュリティ層	システム全体のセキュリティを確保する	ログ整備、トレーサビリティ、権限管理

- プロンプトやスキルを利用して、AIと壁打ちをしながら、自プロダクトのAIセーフティー評価を行う
- 各事業者のユースケースにカスタマイズして活用し実践していく方法のプロンプト・スキルの例を提示

プロンプト例 1 各評価観点のリスクとチェック項目の整理

読み込んだガイド（3 章・4 章）の内容に基づいて、以下のプロダクトのユースケースにおいて各評価観点でのリスクや確認すべきチェック項目を教えてください。

【プロダクト概要】

- プロダクト種別：一般消費者向け（BtoC）健康管理アプリ
- 対象ユーザー：20～60 代の一般生活者
- 主な機能：日常の体調をテキストで入力すると構造化されて管理できる
- 利用データ：ユーザーが入力するテキスト（個人名・ID は取得しない）
- 利用 LLM：API 経由で外部 LLM を利用

以下の形式で回答してください：

1. 各観点における本ユースケース特有のリスク
2. 各観点を最低限確認すべきチェック項目（3～5 個）

- 評価観点ガイドの改訂を軸に、①ガイドの改訂 ②実証 ③出口戦略 ④広報浸透 の4本柱で推進

2025年度に整備した「ヘルスケア領域におけるAIセーフティ評価観点ガイド v1.0」を、AI提供者目線とAI利用者目線の両輪で深化。技術トレンドの取り込み、実証、出口戦略、広報浸透を組み合わせ、社会実装の地盤を一段引き上げる。

① 評価観点ガイド改訂



AI提供者観点

- リスクや課題の整理
- AIエージェント等の動向

AI利用者観点

- 医療機関・生活者の課題

▶ 詳細：活動①-1～①-3

② 実証プロジェクト



- ガイドの有用性検証
- 定量評価
- プロンプト・スキル開発
- ガイドへの反映

▶ 詳細：活動②

③ 出口戦略の検討



- ガイドの中長期の在り方検討
- 海外動向や他領域での制度事例調査
- 適合性評価SWGとの連携

▶ 詳細：活動③

④ 広報・浸透活動



- 学会発表
- 他業界団体との共催イベント
- AISI主催ワークショップ

▶ 詳細：活動④

中長期 日本＝「ヘルスケア×AIセーフティ」の代表事例として、世界的認知の獲得を目指す

活動①-1 一般生活者向けアンケート

- 事業者が想定するリスクと、生活者の実体験リスクのズレ・盲点を検証
- 普及啓発の必要性を定量化し、ガイド改善のエビデンスを収集



検証のポイント

- 10観点と実生活のリスクの乖離
- 注意喚起の設計と利用者リテラシー
- 継続的改善のエビデンス収集



対象者・規模

スクリーニング調査

30,000

ss/7問

全国（性・年代で割付）



本調査

1,030

ss程度/15問

直近1年の健康目的AI利用者（汎用AI利用者とヘルスケア特化型AI利用者を半数ずつ集計）

対象プロダクト

- 汎用生成AI
- ヘルスケア特化型AI

※SaMDは対象外

主な設問

利用実態

- AIの利用頻度・場面
- 体調不良時のファーストタッチ
- 相談した対象者

選択・入力

- AIを選んだ理由
- 専門AIとの誤認
- AIに入力した情報

受け止め・行動

- AI回答の参考度
- 行動への影響
- 回答の検証方法

リスク・不安

- 実害・ヒヤリハット
- 利用時の不安
- 注意喚起の認知

信頼・今後

- 信頼性の判断基準
- 事業者に必要な説明
- 今後の利用意向

活動①-2 医療機関向けアンケート

- AIを導入・運用する「利用者」としての医療機関の実態・課題を把握
- 選定基準や運用上の留意点、事業者へ確認すべき事項を整理



検証のポイント

- AI活用とリスク認識の乖離 
- 現場の「違和感」の抽出 
- 院内ルール・事業者連携の実態 
- 効果的な啓発の対象・形式 

対象者・規模

本調査

300

機関程度／20問程度

病院/診療所に所属する医師個人

主な設問

回答者属性

- 機関種別・病床機能
- 診療科・勤務年数

AI利用実態

- AI利用実態・有用性
- シャドウAIの利用状況

導入・運用

- 導入時の考慮事項
- 院内ガバナンス
- 人の関与（HITL）

インシデント

- 不安や違和感
- インシデント経験
- ベンダーとの連携

事業者への要望

- 開示してほしい情報
- 情報提供の形式・内容
- 院内教育の枠組み

活動①-3 AI提供者向けヒアリング

- Non-SaMDのLLM活用プロダクトを開発・提供する事業者のAIセーフティに係る取組状況を把握
- チェックリストの精緻化や体制構築支援など、ガイド改訂に反映



🔍 把握したいポイント

- 10観点の実装度合いと、評価が難しい観点
- 開発5フェーズのうち課題が集中する箇所
- 医療専門家・法務・セキュリティ人材の確保
- 安全性情報の開示粒度・注意喚起の実態
- 拠り所とする基準・情報源



🏢 対象企業

6 社程度

製薬企業・ヘルスケアテック企業など、規模・業種の多様性を踏まえて選定

ヒアリング項目

① 現状・体制



- 開発体制・リスク感度
- 開発のボトルネック
- 最新技術リスクの把握

② ガイド観点での実践



- 10観点の実装実態と課題
- 開発5フェーズの実践状況

③ ガイドへの意見



- ガイドの認知・活用状況
- 改訂・普及啓発への要望
- 出口戦略への意見

活動② 実証プロジェクト

- 「ヘルスケア領域におけるAIセーフティ評価観点ガイド」の有用性（使いやすさ・妥当性・実効性）を検証し、改善案とベストプラクティスを整理
- また、既存の認証制度を調査し、今後のガイドの方向性を検討



ガイドの有用性検証

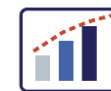
ガイド有用性検証 (定性項目の観点)

- チェックリストの使いやすさ
・改善点
- 観点の抜け漏れ検証
- 体制・ガバナンス・モニタリングの妥当性



ガイド有用性検証 (定量項目の観点)

- 評価項目・サンプルデータセット作成
- シミュレーション評価でプロダクト出力を定量評価
- ガイドの妥当性検証・ベストプラクティス



ガイドの利便性向上

プロンプト／スキル開発

- ガイド活用を支援するプロンプト開発
- エージェントスキルとして試作・試行
- 現場での活用しやすさを向上



活動③ 出口戦略案の検討

- ガイドに実効性を持たせる「出口」として、4つの選択肢を比較検討中
- SWGでの議論を踏まえ、選択肢を再整理する予定



自主的・即時的／今すぐ・低コスト

公的・制度的／導入に時間・コスト大



① AISI・関係省庁等による認定施策

AISI・関係省庁等が関与する形で、公的な認定を付与



② 公的インセンティブへの組み込み

補助金等の給付要件など既存の政策への組み込み



③ 既存制度での参照・要件化

既存制度に推奨事項として位置づけ、段階的に要件化



④ 自己宣言

各社が対応チェックポイントを自社サイト等で公開



SWGで挙げた主な意見



「安全性の確保には相応の投資が必要。その投資に見合うインセンティブがなければ、ガイドの浸透は難しい」



「信頼できることが分かりやすい形で示されるほど、生活者や医療機関がAIを選択しやすくなる」

活動④ 普及広報活動

- ガイドの認知向上と現場での活用促進に向け、11月に医療系学会2件で登壇
- 医療安全・医療情報の実務者に、ガイドの概要と実践方法を発信



ワークショップ

第21回 医療の質・安全学会 学術集会

医療機関におけるTrustworthy AIの実現
～患者安全の視点でAIを選ぶ・使う・管理する～

日時 11月8日（日）13:50～15:20

会場 アクリエひめじ



第1部 インプット 約50分

SWG紹介・調査結果からの課題感
ガイド概要と患者安全視点の活用提案
Trustworthy AIの実現に向けて

第2部 アウトプット 約35分

参加者との意見交換・ディスカッション

パネルディスカッション

第46回 医療情報学連合大会

AISIヘルスケアSWGによるTrustworthy AIの実現に
向けた取組み

日時 11月14日（土）9:10～11:10

会場 札幌コンベンションセンター

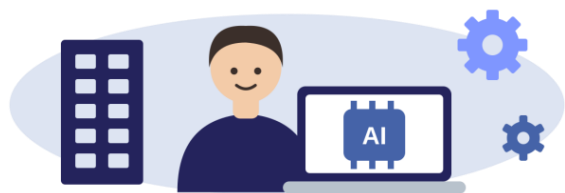


パネルディスカッションの論点

- ① ヘルスケアAIに固有のリスクとは何か
- ② AISIガイドは現場でどう使えるか
- ③ Trustworthy AI実現のために今何が足りないか

今後の活動方針（予定）

- 技術革新を踏まえたガイドのアップデート論点・方針の検討やAIセーフティ評価の検証活動と、ガイドの広報・浸透活動に取り組んでいきます。



2025年度

- Non-SaMDを対象に検討対象とする代表的ユースケースの選定
- 生成AI利活用のリスク構造の明確化、セーフティ評価の観点・シナリオの設計
- 評価ガイド・データセット・ツール等検討



2026年度

- AI開発・提供事業者による評価ガイド・データセット・ツール等の開発・効果検証
- 例：複数の医療機関で導入される生成AIプロダクト・サービスにおいて試行的にAIセーフティ評価を実施



2027年度

- 評価ガイド・データセット・ツール等のヘルスケア業界での広報・浸透活動
- 評価ガイド・データセット・ツール等の継続的アップデート

AISI

Japan AI Safety Institute