

国際医学団体協議会 (CIOMS)

# 医薬品安全性監視における人工知能

---

CIOMS Working Group XIV 報告書



ジュネーブ 2025



国際医学団体協議会 (CIOMS)

## 医薬品安全性監視における人工知能

---

CIOMS Working Group XIV報告書



ジュネーブ 2025

# 日本語版序文

本書は、国際医学団体協議会（Council for International Organizations of Medical Sciences：CIOMS）Working Group XIV による報告書 Artificial Intelligence in Pharmacovigilance の日本語版です。人工知能（AI）を活用した医薬品安全性監視（PV）に関する基本概念、主要な原則、および実務上の論点を、日本語読者がより理解しやすい形で参照できるようにすることを目的として作成しました。

日本語版の作成にあたっては、原著の構成、趣旨、および用語上の整合性を尊重し、内容を要約したり解釈的に変更したりすることなく、できる限り忠実に訳出するよう努めました。また、図表、見出し、用語および参照情報についても、原著との対応関係が維持されるよう配慮しました。

本日本語版の作成にあたり、CIOMS事務局ならびに本報告書を作成された Working Group XIV の皆様に心より敬意を表するとともに、丁寧なご助言とご協力に深く感謝申し上げます。

本日本語版が、日本における医薬品安全性監視実務、研究、教育、および関連する議論の一助となれば幸いです。なお、本書の正式な原典は英語版であり、解釈上の疑義が生じた場合には、原著英語版をご参照ください。

## 訳者

集 弘就

株式会社ヒロファーマコンサルティング® 代表取締役

Hirotsugu Atsumaru

Representative Director /CEO, HiroPharmaConsulting® Co., Ltd.

## 監訳者

栗原 千絵子

神奈川歯科大学 特任教授／日本製薬医学会 前国際製薬医学会代議員

芹生 卓

APCER Life Sciences 取締役会シニアアドバイザー

株式会社セリウムメディカルコンサルティング 代表取締役

日本製薬医学会理事 Medical Safety部会 前部会長

松山 琴音

国立成育医療研究センター臨床研究センター センター長

日本製薬医学会理事 国際製薬医学会次期会長

Chieko Kurihara

Kanagawa Dental University / Former Delegate of JAPhMed, IFAPP

Taku Seriu

APCER Life Sciences Senior Advisor to the Board

Seriu Medical Consulting Ltd. Representative Director

Former Chair of Medical Safety Committee, Director, Board Member, JAPhMed

Kotone Matsuyama

Director, Center for Clinical Research and Development, National Center for Child Health and Development

President-Elect of IFAPP, Director, Board Member, JAPhMed

JAPhMed=The Japanese Association of Pharmaceutical Medicine

IFAPP=The International Federation of Associations of Pharmaceutical Physicians and Pharmaceutical Medicine

## 著作権 © 2025 by the Council for International Organization (CIOMS)

DOI number <https://doi.org/10.56759/cdob6397>

ライセンス: CC BY-NC-SA 4.0.

### ISBN 978-929036110-7

一部の権利は留保されています。本著作物は、Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (<https://creativecommons.org/licenses/by-nc-sa/4.0/>)の下でライセンスされています。

本ライセンスの条件に従い、適切に引用することを条件として、非営利目的で本著作物を複製、再配布、改変することができます。改変を行う場合、改変後の著作物は同一または同等のCreative Commonsライセンスに基づいて利用を許諾するものとします。本著作物の翻訳を作成する場合は、推奨引用に加えて、次の免責事項を付してください。「本翻訳はCouncil for International Organizations of Medical Sciences (CIOMS)により作成されたものではありません。CIOMSは、本翻訳の内容または正確性について責任を負いません。原文（英語版）が正本であり、拘束力を有するものとします。」

第三者資料：表、図、画像等、第三者に帰属する本著作物中の資料を再利用する場合、許可が必要か否かを判断し、著作権者から許可を得る責任は利用者にあります。第三者に帰属する要素の侵害から生じる請求のリスクは、利用者が負うものとします。

**推奨引用:** Artificial intelligence in pharmacovigilance. CIOMS Working Group XIV report. Geneva, Switzerland: Council for International Organizations of Medical Sciences (CIOMS), 2025.

無断転載を禁じます。CIOMSの出版物は、CIOMSウェブサイト (<https://cioms.ch/publications/>) の出版物モジュールを通じてCIOMSから直接入手できます。

詳細は、CIOMS, P.O. Box 2100, CH-1211 Geneva 2, Switzerland, [www.cioms.ch](http://www.cioms.ch), e-mail: [info@cioms.ch](mailto:info@cioms.ch).

**免責事項:** 本書に記載された見解は執筆者に帰属し、必ずしも各執筆者の所属機関や企業の決定、方針、見解を表すものではありません。

デザイン: Agence Gardeners, Annecy (France).

[訳注：上記は原本の著作権に関する記載の日本語訳。本翻訳はCIOMSとの契約に基づくものであり日本語訳についてのCIOMSの免責事項は原本の記載のとおり。本電子版は印刷版作成前の公表版 (Preprint) であり、印刷版作成時に一部修正される可能性がある。日本語訳の最終版は印刷版刊行時に同じwebsiteで入れ替えて公表される。日本語訳の引用は以下（最終版においては「Preprint」を削除する）。

集 弘就, 訳. 栗原 千絵子, 芹生 卓, 松山 琴音, 監訳. 医薬品安全性監視における人工知能 – CIOMS作業部会XIV報告書. スイス・ジュネーヴ. 国際医学団体協議会 (CIOMS). 臨床評価. 2026 ; 54 Suppl 43. Preprint. ]

# 謝辞

国際医学団体協議会（Council for International Organizations of Medical Sciences：CIOMS）は、本報告書の作成にあたり、ICH設立メンバーを含む学界、産業界、国際および国内の医薬品安全性監視（PV）組織、医薬品規制当局の専門家の皆さまが、貴重な時間と豊富な経験を提供してくださったことに感謝します。

会議を主導し積極的な役割を果たしてくださったメンバーに感謝するとともに、報告書の議論、起草、再起草、レビューにおいて寄せられたあらゆる意見を高く評価します。特に、章およびセクションのリーダーと共同リーダーに感謝します：Hua Carroll、Julie Durand、Manfred Hauben、Thomas Henn、Satoko Hirokawa-Voorburg、Vijay Kara、Benny Ling、Denny Lorenz、Elizabeth MacEntee Pileggi、Niklas Noren、Ravi Patel、Walter Straus。

作業部会（WG）XIVの編集チームについては、チームの作業調整を担ったElizabeth MacEntee Pileggiへの特別な謝意とともに言及する価値があります。CIOMSは、編集チームの他のメンバーにも感謝します：Justyna Amelio、Andrew Bate、Taxiarchis Botsis、Hua Carroll、Julie Durand、Manfred Hauben、Thomas Henn、Vijay Kara、Benny Ling、Denny Lorenz、Lembit Rago、Monica Da Luz Carvalho Soares、Niklas Noren、Ravi Patel、Walter Straus。さらに、報告書を校正したCheryl Renzに深く感謝します。

WGには、Douglas Domalikをリーダーとする用語集チームがあり、本分野の多くの用語の定義に貢献し、報告書を幅広い読者にとって利用しやすいものとなりました。用語集チームのメンバーは、Neal Grabowski、Thomas Henn、Vijay Kara、Benny Ling、Manuela Messelhauserです。

WGは、PVにおける人工知能（AI）の適用可能性を示すため、一連のユースケース（付録3参照）の開発に多くの時間を費やしました。CIOMSは、Vijay Kara、Hans-Jorg Rödmring、Taxiarchis Botsis、Adrian Berridge、Manfred Haubenに感謝します。

CIOMSは、2025年に約6週間にわたりCIOMSウェブサイト上で公開された報告書案のパブリックコンサルテーションに参加し、WGの広範なネットワークを通じてコメントを寄せてくださった多くの方々（付録6参照）に感謝します。寄せられたすべてのコメントと提案は、ありがたく受け取り、検討しました。

WGは多数の対面会合を行い、CIOMSは、そのうち2回の会合を親切に主催いただいた武田薬品工業株式会社およびMerck KGaAの寛大なご支援を歓迎します。

最後に、思慮深く時間を割き、重要な役割において専門知識を提供くださったものの、事情により氏名を公表できないメンバーの皆さまに丁寧にお礼を申し上げます。

CIOMSでは、Hervé Le LouëtとLembit RagoがWG会合の議長を分担し、Lembit RagoSanna Hill、Kateriina Rannula、Sue le Rouxがプロジェクトを管理しました。

スイス、ジュネーブ、2025年  
Lembit Rago M.D., Ph.D.  
CIOMS事務局長

# 目次

日本語版序文.....	II
謝辞.....	IV
略語一覧.....	X
序文.....	XIII
エグゼクティブ・サマリー.....	1
<b>第1章 序章.....</b>	<b>5</b>
1.1. 適用範囲.....	10
<b>第2章 ランドスケープ分析.....</b>	<b>13</b>
2.1. これまでの医薬品安全性監視における人工知能の利用.....	13
2.2. 規制上の考慮事項.....	16
<b>第3章 リスクに基づくアプローチ.....</b>	<b>27</b>
3.1. はじめに.....	27
3.2. リスクアセスメント.....	30
3.3. 問題の検出とリスクの軽減.....	33
3.4. リスクに基づくアプローチのレビューと文書化.....	34
<b>第4章 ヒューマン・オーバーサイト.....</b>	<b>37</b>
4.1. はじめに.....	37
4.2. 人間の関与と監督に関する考慮事項.....	38
4.3. 従来の役割の変革.....	40
<b>第5章 妥当性と頑健性.....</b>	<b>43</b>
5.1. はじめに.....	43
5.2. 仕様・設計.....	44
5.3. 性能評価.....	46
5.4. ヒューマン・イン・ザ・ループによる人工知能システムの評価.....	50
5.5. 継続的インテグレーションとデプロイ.....	51
<b>第6章 透明性.....</b>	<b>55</b>
6.1. はじめに.....	55
6.2. 人工知能利用の開示.....	56
6.3. 人工知能モデルに関する透明性.....	56
6.4. 説明可能性.....	58
6.5. パフォーマンスに関する透明性.....	61

<b>第7章 データプライバシー</b> .....	<b>65</b>
7.1. はじめに .....	65
7.2. 倫理的配慮 .....	66
7.3. 実際の配慮 データプライバシーのサポート .....	72
7.4. 結論 .....	76
<b>第8章 公正性と衡平性</b> .....	<b>81</b>
8.1. はじめに .....	81
8.2. 公正性と衡平性の考慮事項と医薬品安全性監視 .....	82
8.3. 公正性と衡平性に対する潜在的脅威の要因 .....	83
8.4. リスク、影響、緩和策 .....	86
8.5. 主な軽減戦略 .....	87
<b>第9章 ガバナンスとアカウントビリティ</b> .....	<b>89</b>
9.1. はじめに .....	89
9.2. ガバナンスフレームワーク .....	90
9.3. トレーサビリティとバージョン管理 .....	95
<b>第10章 医薬品安全性監視における人工知能の開発と展開に関する 今後の検討</b> .....	<b>99</b>
10.1. 医薬品安全性監視における人工知能の進化と将来 .....	99
10.2. 医薬品安全性監視の長期的・将来的な変革的役割：予測から検出・予防へ .....	100
10.3. AIの今後の開発と展開および指針原則 .....	101
10.4. 医薬品安全性監視における人工知能の開発と展開に関する今後の検討の結論 .....	105
<b>付録1. 用語集</b> .....	<b>107</b>
<b>付録2. 指針原則の比較表</b> .....	<b>123</b>
<b>付録3. ユースケース</b> .....	<b>135</b>
ユースケースA: ケース処理のための大規模言語モデルデータ抽出 .....	136
ユースケースB: ケースの重複除外 .....	140
ユースケースC: 人工知能翻訳アシスタント .....	144
ユースケースD: コンテキスト対応構造化照会言語のための大規模言語モデル .....	147
ユースケースE: 薬物有害反応の因果関係評価 .....	150
ユースケースF: シグナル検出を支援するプロセス効率 .....	154
ユースケースG: 生成AI: 医薬品安全性監視評価を促進するための大規模な非構造化 安全性文書リポジトリからの合成と要約 .....	158
ユースケースH: (ヒドロキシ)クロロキン網膜症の診断と予測を支援する人工知能 .....	163

<b>付録4. 説明可能性ならびに公正性と衡平性に関するコンテンツ</b> .....	<b>169</b>
説明可能性に関する具体的事例 .....	169
公正性と衡平性に関するコンテンツ .....	174
<b>付録5. CIOMS 作業部会のメンバーと会合</b> .....	<b>177</b>
<b>付録6. パブリック・コンサルテーションのコメンテーター</b> .....	<b>181</b>

## 図一覧

図 1: 代表的なシグナル管理プロセス .....	7
図 2: 医薬品とワクチンの有害事象報告に関する世界保健機関の グローバルデータベースであるVigiBaseの経時的成長 .....	8
図 3: FDA有害事象報告システム (FAERS) データベースの経時的成長 .....	9
図 4: モデルのリスクマトリックス .....	32
図 5: 大規模言語モデルのプライバシー保護に関する研究の状態 (2025年6月現在) .....	76
図 6: 条件付き推論木を用いたコントロールの自動選択と偽陽性シグナルの棄却の 実装をまとめたフローチャート .....	154
図 7: 初期の人工知能アーキテクチャの概要 .....	159
図 8: 米国FDAの情報可視化プラットフォームのユーザーインターフェイス： アクセシビリティの分類に積極的に貢献する機能に焦点を当て システムの機能を表示 .....	170

## 表一覧

表 1: バブリックドメインで記述された医薬品安全性監視における 人工知能ソリューションの導入例 .....	16
表 2: 地域および国の政府機関、国際機関におけるAI指針原則の比較 - CIOMS Working Group XIV作成 .....	17
表 3: ステークホルダーに開示する人工知能モデルの主要な側面 .....	57
表 4: 人工知能モデルの推定パフォーマンスに関する透明性を確保するために開示する 関連する側面 .....	61
表 5: ブラジルにおけるデータの二次利用に関するデータプライバシー規制 .....	69
表 6: 中国におけるデータの二次利用に関するデータプライバシー規制 .....	70
表 7: ドイツにおけるデータの二次利用に関するデータプライバシー規制 .....	71
表 8: 日本におけるデータの二次利用に関するデータプライバシー規制 .....	72
表 9: ガバナンスフレームワークのグリッド .....	92
表 10: CIOMS WORKING GROUP XIVの地域および国の政府機関、国際機関における 人工知能の指針となる原則の比較 - 原則の抜粋 .....	123
表 11: ユースケースA: ガバナンスフレームワークとの整合性 (詳細) .....	138
表 12: ユースケースB: ガバナンスフレームワークとの整合性 (詳細) .....	142
表 13: ユースケースC: ガバナンスフレームワークとの整合性 (詳細) .....	145
表 14: ユースケースD: ガバナンスフレームワークとの整合性 (詳細) .....	148
表 15: ユースケースE: データのモデリング .....	150
表 16: ユースケースE: ガバナンスフレームワークとの整合性 (詳細) .....	152
表 17: ユースケースF: ガバナンスフレームワークとの整合性 (詳細) .....	156
表 18: ユースケースG: ガバナンスフレームワークとの整合性 (詳細) .....	161
表 19: ユースケースH: ガバナンスフレームワークとの整合性 (詳細) .....	165

## 略語一覧

ADR	医薬品副作用 (Adverse Drug Reaction)
AE	有害事象 (Adverse Event)
AERS	有害事象報告システム (Adverse Event Reporting Systems)
AI	人工知能 (Artificial Intelligence)
AIA	アルゴリズム影響評価 (Algorithmic Impact Assessment)
AIDA	人工知能およびデータ法 (Artificial Intelligence and Data Act、カナダ)
AI-MD	人工知能医療機器 (Artificial Intelligence Medical Device)
AKI	急性腎障害 (Acute Kidney Injury)
ATC	解剖学的治療化学分類 (Anatomical Therapeutic Chemical)
BCR	拘束的企業準則 (Binding Corporate Rules)
BLEU	バイリンガル評価アンダースタディ (Bilingual Evaluation Understudy)
CDC	疾病対策センター (Centers for Disease Control and Prevention)
CGRP	カルシトニン遺伝子関連ペプチド (Calcitonin Gene-Related Peptide)
CIOMS	国際医学団体協議会 (Council for International Organizations of Medical Sciences)
CNN	畳み込みニューラルネットワーク (Convolutional Neural Networks)
COU	使用目的 (Context of Use)
CSV	コンピュータ化システムバリデーション (Computerized System Validation)
EDSTP	新興医薬品安全性技術プログラム (Emerging Drug Safety Technology Program)
EEA	欧州経済領域 (European Economic Area)
EHR	電子健康記録 (Electronic Health Records)
EMA	欧州医薬品庁 (European Medicines Agency)
ETHER	健康電子記録のイベントベーステキストマイニング (Event-based Text-mining of Health Electronic Records)
EU	欧州連合 (European Union)
EU AI Act	欧州人工知能法 (European Artificial Intelligence Act)
FAERS	FDA有害事象報告システム (Food and Drug Administration Adverse Event Reporting System米国)
GAMP 5	適正自動化製造実践ガイド第5版 (Good Automated Manufacturing Practice 5)
GDPR	一般データ保護規則 (General Data Protection Regulation)
GenAI	生成型人工知能 (Generative Artificial Intelligence)
GPT	生成事前学習トランスフォーマー (Generative Pre-trained Transformer)
GVP	製造販売後安全管理の基準 (Good Pharmacovigilance Practices)

GxP	適正 [x] 実施基準 (Good [x] Practices)
HCP	医療専門職 (Healthcare Professional)
HFE	ヒューマンファクターおよび人間工学 (Human Factors and Ergonomics)
HIC	ヒューマン・イン・コマンド (人間による統括管理) (Human-in-Command)
HIPAA	医療保険の携行性と責任に関する法律 (Health Insurance Portability and Accountability Act)
HITL	ヒューマン・イン・ザ・ループ (Human-in-the-Loop)
HOTL	ヒューマン・オン・ザ・ループ (Human-on-the-Loop)
ICH	医薬品規制調和国際会議 (The International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use)
ICSR	個別症例安全性報告 (Individual Case Safety Report)
InfoViP	情報可視化プラットフォーム (Information Visualization Platform)
IRB	機関倫理審査委員会 (Institutional Review Board)
ISIC	国際皮膚画像共同研究 (International Skin Imaging Collaboration)
IT	情報技術 (Information Technology)
KPI	重要業績評価指標 (Key Performance Indicator)
LIME	局所的解釈可能モデル非依存説明 (Local Interpretable Model-Agnostic Explanations)
LLM	大規模言語モデル (Large Language Models)
LRS	網膜症可能性スコア (Likelihood of Retinopathy Score)
MAH	販売承認保有者 (Marketing Authorisation Holder)
MedDRA	医薬品規制活動用語集 (Medical Dictionary for Regulatory Activities)
mfERG	多局所網膜電図 (Multifocal Electroretinography)
MHRA	医薬品・医療製品規制庁 (Medicines and Healthcare products Regulatory Agency、英国)
ML	機械学習 (Machine Learning)
ML-DSF	機械学習対応医療機器ソフトウェア機能 (Machine Learning-Enabled Device Software Functions)
NLP	自然言語処理 (Natural Language Processing)
NLQ	自然言語クエリ (Natural Language Query)
OCT	光干渉断層計 (Optical Coherence Tomography)
OECD	経済協力開発機構 (Organisation for Economic Co-operation and Development)
OTC	一般用医薬品 (Over the Counter)
PASS	製造販売後安全性試験 (Post-Approval Safety Studies)
PD	薬力学 (Pharmacodynamic)
PDMP	処方薬監視プログラム (Prescription Drug Monitoring Program)
PHI	保護対象医療情報 (Protected Health Information)



# 序文

国際医学団体協議会（CIOMS）は、医薬品の開発および安全性に関する倫理的ならびに科学的側面に対処するガイドラインを策定することにより、現代の医薬品安全性監視（pharmacovigilance: PV）の発展において重要な役割を果たしてきました。特にCIOMSは、PVにおけるシグナル検出の実務的側面に関する指針に加え、薬物有害反応（ADR）の収集および報告を体系化するための枠組みを支援する文書を公表し、医薬品安全性モニタリングにおける国際協力および標準化の推進に貢献してきました。

1950年代後半から1960年代初頭にかけて発生したサリドマイド薬害は、世界的な医薬品安全性実務における重大な欠陥を明らかにし、体系的なデータ収集および国際的調和の必要性を強く認識させる契機となりました。これを受け、世界保健機関（WHO）は1968年に国際医薬品モニタリングプログラムを設立し、各国間で個別症例安全性報告を共有し、データ実務の調和を図る取組みを開始しました。これらの基盤的活動を背景として1980年代後半から1990年代にかけて、CIOMS Working Group Iによる「医薬品副作用のモニタリングと評価」（1985年）および「医薬品副作用の国際報告」（1987年）、さらにCIOMS Working Group Vによる「医薬品安全性監視における現在の課題：実務的アプローチ」（1999年）など、重要なCIOMS報告書が公表されました。その後のCIOMS Working Groupによる継続的な活動および「ヒト用医薬品の技術要件の調和のための国際会議（ICH）」の設立は、従来指摘されてきた医薬品安全性への断片的アプローチに対処することを目的とし、シグナル検出プロセスのみならず、有害事象（AE）の収集および報告を標準化する国際的枠組みの形成に寄与しました。

技術の進歩はPVの在り方を大きく変化させています。スマートフォンやウェブアプリケーションを通じた直接的な有害事象報告から、<sup>1</sup>大規模医療ネットワークを介したデータ統合に至るまで、多様な手段により、有害事象をほぼリアルタイムでプロトコルに基づき評価することが可能になりつつあります。<sup>2</sup>このような背景のもと、人工知能（artificial intelligence: AI）のPVへの応用は、PV活動の効率性と品質の双方を向上させる可能性を有していますがその利用には人間による意思決定が不可欠であり、規制上の期待に適合することが求められます。そのため、説明可能で、信頼性があり、責任あるAIの利用が必要とされます。同時に、倫理的かつ責任あるデータ利用も極めて重要であり、これはAI成果の完全性を支え、これらの技術に対する社会的信頼の形成に寄与します。

AIが進化し続け、生物医学研究に広範な影響を及ぼすにつれて、PV実務への統合および影響の拡大は避けられないものとなっています。AIがもたらし得る変革的可能性を踏まえると、この分野におけるAIの開発、検証および展開の在り方について、厳密かつ前向きな議論を行うことが不可欠です。

CIOMSは2022年初頭、「PVにおけるAI」を検討するWorking Group XIVを設立しました。それ以降、生成AI（GenAI）の急速な進展および広範な利用可能性により、本分野は著しい進歩を遂げています。PVにおけるGenAI活用への関心が高まる一方で、本Working Groupは、高度に規制された領域であるPVに特有の課題を踏まえた「適切な利用」に焦点を当てる必要性を認識しており、可能かつ有益な場合には、AI一般に関する議論と区別して整理することを目指しました。

<sup>1</sup> 医薬品

本報告書において「医薬品」とは、疾病の治療、予防または診断に使用される製品、ならびに身体の機能を回復、補正または調整する一部の製品を指します。本報告書では、これらは国および地域の医薬品規制当局の活動範囲に属する製品を意味します。ワクチンおよび医薬品・機器の組み合わせ製品も本報告書における「医薬品」の定義に含まれます。「医薬品」と同義で使用される用語として、drugs（薬物）、medications（薬剤）、medicinal products（医薬品）があります。

出典：CIOMS Working Group XI

本報告書は、PVにおけるAIの開発および利用に関する原則およびグッドプラクティスの一般的枠組みを提示することを目的としています。技術的詳細に関する指示を与えることよりも、AI能力の進展に伴っても持続的に関連性を保ち得る原則を示すことを重視しています。また、本報告書は、AI一般の利用に関する広範な論点ではなく、PVに特有、または特に関連性の高い応用に焦点を当てています。

本報告書は、規制当局、産業界、学術研究者、臨床医、患者、倫理専門家、技術ベンダーおよび国際機関など、PVにおけるAIソリューションの開発および利用に関心を有する幅広い関係者に対し、指針を提供することを目的としています。

## 序文 – 参考文献

- 1 Liyanage PH, Madhushika MT, Liyanage PLGC. Effectiveness of mobile applications in enhancing adverse drug reaction reporting: A systematic review. BMC Digit Health, 2025;3:15. <https://doi.org/10.1186/s44247-025-00153-9> (Journal full text)
- 2 Desai RJ, Matheny ME, Johnson K, Marsolo K, et al. (2021). Broadening the reach of the FDA Sentinel system: A roadmap for integrating electronic health record data in a causal analysis framework. NPJ Digit Med. 2021;4:170. <https://doi.org/10.1038/s41746-021-00542-0> (Journal full text)

# エグゼクティブ・サマリー

本報告書「医薬品安全性監視における人工知能」は、医薬品安全性監視（PV）、コンピュータサイエンス、数学、規制、法律、医学、人権、心理学および社会科学が交差する、急速に発展しつつある学際領域を扱っています。したがって、医薬品の場合と同様に、PVにおける人工知能（AI）の利用について、適応、用法・用量、副作用、警告および使用上の注意に相当する事項を確立することが重要です。これらは、研究および実務実装を効果的かつ安全で責任ある形で推進するため、多様な背景を持つ人々によって明確に定義され理解される必要があります。対象には、バイオ医薬品産業、規制当局、学術分野においてPVに従事する専門家、研究者および意思決定者が含まれます。また、シグナル管理および個別症例安全性報告（ICSR）処理のあらゆる側面を含むPV向けAIソリューションを開発するソフトウェアベンダーも含まれます。本報告書は、応用科学研究や公的議論への参加あるいは組織内での評価および意思決定を通じて本領域に積極的に関与するために必要な用語および概念理解を提供します。

他のCIOMS報告書の主題と比較しても、PVにおけるAIの潜在的危険性および関連事項は本報告書の主要成果の中核的要素となっています。急速に進化し高度化し、しばしば不透明な技術は、いわゆる技術の「ハイプサイクル」に見られるように、実際の有用性とかげ離れた過度な期待や過大評価を招く可能性があるためです。一方で、人間保護を目的として複数の組織により策定された中核的指針が安全網として存在しており、PVにおけるAIはその枠組みの中で発展していく必要があります。本報告書は、それらの指針および策定主体を示します。本報告書の大部分は、リスクに基づくアプローチ、人間による監督、妥当性と頑健性、透明性、データプライバシー、公正性と衡平性（fairness and equity）、ガバナンスと説明責任という原則により構成されています。これらの指針における重要な考慮事項は本文全体で詳述されており、以下に要約されます。

これまでのCIOMS報告書と同様に、本報告書も規制当局、学術界、産業界を含む複数のステークホルダーによるコンセンサスに基づいて作成されています。ワーキンググループ（WG）は、AI分野の進展速度が極めて速く、規範的な詳細規定は短期間で陳腐化する可能性が高いことを認識していました。そのためWGは、今後数年間にわたりPV専門家にとって有用であり続ける共通原則に焦点を当てることを決定しました。PVは、現代社会の多くの側面を変革しつつある数多くのAI応用の一分野に過ぎません。本報告書はまた、各国政府によるAIへの関心の高まりの恩恵も受けています。複数の政府は、医薬品開発におけるAIのみならず、より広範なAI利用全般に関する法制度およびガイダンスを公表しています。

**リスクに基づくアプローチ** AIをPVプロセスに統合するには、AIシステムに内在する潜在的な不正確性および変動性、ならびに個人および社会の安全と福祉への影響を考慮する必要があります。第一に、誤りが人間に対して深刻な悪影響を招き得るような重大な影響を伴うもの（high-stakes）かどうか、という点です。第二に、AIが監督なしの単独モードで使用されるのか、それとも人間とコンピュータの相互作用の中で使用されるのか、という点です。リスクに見合ったヒューマン・オーバーサイトを組み込むことで、組織はAIの能力を最大限に活用しつつ、患者安全およびPVステークホルダーへの悪影響を回避できます。リスクに基づくアプローチは、人間による監督、妥当性と頑健性、透明性、公正性と衡平性、データプライバシーに関する取組みに適用されるべきです。また、リスク評価ではAIシステム自体、利用状況、影響の重大性および発生可能性を考慮し、定期的にレビューおよび文書化する必要があります。

**ヒューマン・オーバーサイト（人間による監視）** ヒューマン・オーバーサイトは、PVにおけるAIの性能最適化を支援し、信頼性および説明責任を高めます。AIソリューションに対するヒューマン・オーバーサイトの程度と性質は、品質保証の原則を組み込んだリスクベースドであるべきです。ヒューマン・オーバーサイトには、人間と機械の相互作用の結果として意思決定がなされる「ヒューマン・イン・ザ・ループ」の場合もあれば、機械が自律的に意思決定を行う、または結果を返し、それを人間が確認する「ヒューマン・オン・ザ・ループ」の場合もあります。ヒューマン・オーバーサイトは、意図された課題に対して目的に適合した性能水準（すなわち妥当性）を定義するために必要です。これには、許容可能な性能ベンチマークの事前定義、現実的な環境におけるモデル開発および試験／バリデーションのための適切なデータの選択、継続的な品質評価プロセス、ならびに必要に応じたモデルの再学習または動的／オンライン学習が含まれます。PVにおける自動化およびAIの利用拡大は、従来の役割と能力を変容させるため、適切な変更管理および教育訓練戦略が必要となります。

**妥当性と頑健性** PVのステークホルダーは、提案されるAIソリューションを継続的かつ批判的に評価することを学ばなければなりません。性能評価は、現実的な条件下で、意図された用途に対して許容可能で頑健な結果を示さなければなりません。そのような評価は、定性的かつ定量的であり、学際的な取組であるべきで、関連する多様な事例にまたがって実施される必要があります。評価では、バイアスの発生可能性を低減し検出し、意図された展開領域全体で十分かつ一般化可能な性能を促進し、ユーザビリティを評価し、性能低下に関連する状況を特定するために、関連するデータ型（例：自発報告、臨床試験、文献などのデータソース）、報告者／患者特性、ならびに多様な医薬品、ワクチンおよび有害事象を十分に代表するデータを用いるべきです。PVにおける多くの応用は、稀な事象またはパターン（例：安全性シグナルや重複症例）の認識に焦点を当てているため、評価対象の結果が十分な頻度で含まれる代表的な試験セットを得るために、強化戦略が必要となる場合があります。性能評価結果がリアルワールドの環境に一般化可能であるように努めるべきです。

**透明性** AIソリューションがいつ、どのように使用されるかを明示することは、ステークホルダーの間で信頼を構築するうえで極めて重要です。中核的なPV業務に導入されるAIソリューションの性質は、それに対応する価値提案とともに伝達されるべきであり、これにはモデル開発およびアーキテクチャ、想定される入力と出力、人間とコンピュータの相互作用の性質が含まれます。AIソリューションの有効性と限界を十分に特徴づけるために、性能評価結果では、参照標準およびサンプリング戦略を含めて、使用した試験セットの範囲と性質を記述すべきです。性能指標は意図された課題に適合しているべきであり、適切なベンチマークとの比較を行い、正しい出力例と誤った出力例の代表例に関する定性的レビューで補完されるべきです。説明可能性は、内部の意思決定経路が非常に複雑かつ非線形で、技術的素養のある者にとっても解明困難な、いわゆる第一種のブラックボックスに該当するモデルに関連する重要な概念です。Explainable AIとは、そのような経路について、ブラックボックスがどのようにして出力に至ったかに関する、もっともらしい仮説を返す一連の技術です。これが可能であることは、モデル構築やトラブルシューティング、信頼の構築、監査可能性および説明責任の確立、自身に不利益を及ぼし得るAIの結果に対して人間が異議を唱える根拠の提供、規制遵守、ならびに科学的仮説の生成に有利となり得ます。可能であれば、AIモデルがどのような一般原理および論理によって機能し、その結果／予測に至るのかの説明を共有すべきであり、そのような説明可能性が欠如している場合には、そのことを認めたくうえで、その含意を議論すべきです。ただし、Explainable AIの手法には限界があり、それらが提供するのはいくらももらしい仮説にすぎず、AIが実際にその仮定された意思決定経路を用いたことを保証するものではありません。

**データプライバシー** PVにおけるAI利用を評価する倫理的枠組みは、人間を対象とする研究活動に関する標準の原則の中に位置づけられます。PVの日常業務におけるAI利用にとって極めて重要な原則の一つは、データプライバシーの不可侵性です。AIを支えるハードウェアおよびソフトウェアの能力が高まるにつれ、大規模で相互連結されたデータベースを構築できる可能性が広がっており、さらに大規模言語モデル（LLM）には患者再識別の潜在的可能性があります。これは導入前のデータ保護影響評価によって対処されます。これらは、データプライバシーを保護する従来の安全策に対して継続的な課題をもたらす可能性があります。この文脈では、AIの開発および導入のワークフロー全体を通じて、非常に機微な個人情報および健康情報が、広範かつ学際的なステークホルダーに開示され得る機会が複数存在します。その結果、各国ではこれらのデータを保護することを目的とした法令およびガイダンスが整備されつつあります。PV専門家およびソフトウェアベンダーなどのステークホルダーは、GenAIによるデータプライバシー侵害リスクの高まりにより、規制遵守を担保するために既存の用いられている手順を再評価する必要があります。

**公正性と衡平性 (Fairness and Equity) \*** PVにおけるAIの公正かつ衡平な利用に関する主要な規制上・倫理上の要請には、公正性と衡平性を支援すること、有害な明示的バイアス、差別、および不正確な結果をモデル開発および導入の過程で伝播または増幅させないこと、ならびに特定のサブ集団に十分なサービスが提供されない状況避けることが含まれます。こうした問題は、そもそもAIソリューションを導入するか否かという初期判断にまで及ぶことがあります。衡平性は、PVにおけるAIが、評価対象となる特定の医薬品に曝露されると想定される集団にとって適切な出力を返すことを保証するための措置を講じることで促進され得ます。可能な場合には、明示的または潜在的なバイアスをスクリーニングし、特定し、除去することが、リスク低減、AIの適用可能性および限界の判断、ならびに許容可能な性能の定義にとって重要です。学習および性能評価に用いる参照データセットは十分な代表性を有しているかについて精査されるべきであり、可能であれば、関連するサブグループごとに性能評価を行うべきです。不十分な参照データは、公正性および衡平性の不十分さの原因となることが少なくありません。

[\*訳注：本翻訳ではequityの訳語の漢字を「公平性」ではなく「衡平性（こうへいせい）」としているが、どちらの漢字を用いた場合も意味は同じである。「平等」（equality）は人間の基本的権利であり、その実現のために等しく機会・資源が配分・提供されるべきとする概念であるのに対して、「公平性」「衡平性」（equity）はニーズに応じて配分・提供され、結果としての基本的人権の平等が達成されるべきという文脈で使われる概念である。訳語の漢字として「公平性」「衡平性」のそれぞれが使用される領域や傾向の違いがあるが、その傾向は固定したものではないためここでは違いについて言及しない。]

**ガバナンスとアカウントビリティ** 強固なガバナンスと明確なアカウントビリティは、AI施策の成功にとって極めて重要です。これらの原則は、AIソリューションが安全に、責任をもって、倫理的に使用され、適用されるすべての法的・規制上の要件を遵守しつつ、ステークホルダー間の信頼と透明性を醸成することを支えます。明確に定義された役割と責任は、すべてのステークホルダーが、自らのアカウントビリティと義務を理解し、AIソリューションを効果的に監督できるようにするために不可欠です。AI技術が進化するにつれて、ガバナンスおよびアカウントビリティの枠組みも適応させていく必要があります。新たなリスクと課題が出現し、それに応じて原則と実務の更新が求められます。こうした変化に先んじるためには、継続的な見直しと適応が不可欠です。これには、前述の指針原則を実務に適用するために提案されたガバナンス枠組みグリッド（第9章「ガバナンスとアカウントビリティ」参照）の適応および改善も含まれます。

**医薬品安全性監視における人工知能の開発と展開に関する今後の検討** 医薬品安全性監視（PV）における人工知能（AI）の導入拡大によりリアルタイムまたは準リアルタイムでのシグナル検出のための迅速なデータ収集、評価および報告が優先され、それに対応することが期待されます。これはまた、データが十分にある状態（warm-start）の予測シナリオからデータが無いか不足している状態（cold-start）の予測シナリオへの相対的な移行（すなわち、市販後から開発初期段階への移行）を伴う可能性があります。これは、例えばプロセスの効率化や、患者プライバシーを含む、より広い医療環境およびその外側にまで及ぶ変化を引き起こすことにより、こうした技術的進歩をどのように活用するかを根本的に変える可能性があります。また、PVにおけるAIの導入は、臨床現場においても拡大し、有害反応の一次二次、および三次予防を支援し得ると考えられます。人間が in-the-loop または on-the-loop にどの程度関与し続けるかは、課題の性質（例：定型化された業務か、専門的な臨床的・科学的判断を要する業務か）によって、前述のリスクベースアプローチと整合する形で決定されますが、一部のAIベースの専門家システムは、最終的には洗練された医学的・科学的判断を発展させる可能性もあります。

本報告書で示した指針原則が、中核的な考慮事項として維持され、具体的な利用文脈において責任をもって適用されることが極めて重要です。これらの原則は、PVおよび医療一般におけるAIの進歩と適用に応じて進化・適応していく必要があります。そのためには、プロセス、データ、ならびにAIの能力と限界についての十分な理解と柔軟性が求められます。これは、PVにおけるAI利用が偏りなく、透明で、安全であり続け、誤用または偶発的危害を防ぐことを確保するためです。適切なヒューマン・オーバーサイトは、規制上および倫理上の安全策を含め、適用される技術的進歩と同様に重要であり続けます。

# 第1章 序章

AIシステムとは、明示的または暗黙的な目的のために、受け取った入力から、予測、コンテンツ、推奨、または意思決定など、物理的または仮想的環境に影響を及ぼし得る出力をどのように生成するかを推論する機械ベースのシステムです。<sup>1</sup> AIシステムは、展開後の自律性および適応性の程度においてさまざまです。この定義には、人間の専門知識に基づいて明示的にプログラムされたシステムから、機械学習 (machine learning: ML) に基づく手法、さらにディープニューラルネットワークのようなより複雑なアプローチまでが含まれます。ただし、AIシステムを利用する組織が、それぞれの業務プロセスにおいて、より限定的な定義を採用する場合があることも、ここで認識しておきます。

医薬品安全性監視 (PV) の文脈においてAIシステムを利用する目的は、PV、患者安全および規制遵守を強化し、最終的には治療のベネフィット・リスク最適化に資する意思決定を支援することにあります。PVは、製薬企業、保健当局、医薬品監視センター、学術機関のみならず、臨床現場においても実践されており、AIはこれらすべての環境においてPVへの応用が進みつつあります。<sup>2</sup>

AIソリューションは、PVにおける特定の目的に対応するよう設計されるべきです。AIソリューション全体は、1つまたは複数のAIシステムによって構成される場合があります。AIシステムには、モデルそのものに加え、ユーザーインターフェイスやデータ処理パイプラインなど、運用に必要な構成要素が含まれます。これらの中核にはAIモデルがあります。AIモデルは、パラメータを用いてデータ内のパターンや関係を学習し、システムが時間の経過とともに適応してモデル性能を向上させたり、知識検索システムを支援したりすることを可能にします。

統計的手法によるシグナル検出などの比較的単純なAIシステムは、すでに数十年にわたりPVで広く利用されてきました。<sup>3</sup> しかし過去10年余りにおいて、特に画像解析および自然言語処理 (natural language processing: NLP) の分野でAI能力は飛躍的に向上しました。これらの進歩によりAI利用は大幅に拡大しました。さらに、計算能力およびモデルアーキテクチャの継続的な進展により、大規模電子データベースの構築および連結が可能となりました。その結果AIはライフサイエンスを含む多くの学問領域へ応用されるようになりました。<sup>4</sup> ライフサイエンス分野では、医用画像診断、創薬および医薬品開発、ゲノミクス、精密医療 (precision medicine)、公衆衛生、医療提供など、多様な領域にAIが適用されています。<sup>5</sup>

AIの進展により、医薬品分野は臨床、規制およびPV実務全体において急速な変革期を迎えつつあります。エンドツーエンドのプロセス合理化を通じて製品開発および市場投入の加速が期待されています。同時に、臨床および市販後安全性管理の強化への関心も高まり、安全性シグナルの早期特定、さらには予測的評価、およびベネフィット・リスク評価の高度化が目指されています。臨床現場では、薬物有害反応の早期診断 (すなわち二次予防および三次予防) を目的としてAIの検証または導入が進められています。例として、ヒドロキシクロロキン網膜症の早期検出、<sup>6</sup> ジゴキシン毒性の検出、<sup>7</sup> パーキンソン病患者における薬剤誘発性運動障害の検出などが挙げられます。<sup>8</sup>

これらの進歩は、大規模に統合されたデータセットおよび帰納的論理を活用するものであり、AIモデルが明示的な規則や人間の介入のみに依存するのではなく、蓄積されたデータを利用して妥当性のある推論を行うことを可能にします。このアプローチは、新規

改良型、または補完的な解決策を提供するAIシステムの開発を促進します。PV領域においてAIを成功させるための重要な要素は、電子医療記録（EHR）、保険請求データベース、レジストリ、モノのインターネット（IoT）、接続デバイスなど、多様なデータソースから得られる品質の異なる大量の異種データを連結し、解析できる能力にあります。

医療データの活用は、新規治療法のより迅速な開発、患者アウトカムの改善、医療費の削減につながる可能性があります。また、従来は特定されなかった可能性のある、新規で有用かつ実行可能なインサイトを引き出す可能性もあります。一方で、外部データセットとの連結は、追加的なプライバシー上の影響およびリスクを伴う場合があります。これらのリスクを軽減するために、プライバシー保護型レコードリンケージ手法を適用することにより、患者機密性を維持しながら、安全かつ倫理的なデータ統合を実現できます。<sup>9,10</sup> したがって、患者安全の向上を支援するうえで、データアクセスの重要性を効果的に伝達する必要性が高まっています。

PVにAIを組み込むためには、その潜在的な利益およびリスクを十分に評価することが必要であり、関係者が既存の実務に対する影響を理解できるようにすることが求められます。技術変化の速度が極めて速いことを踏まえ、本書はPVにおけるAIの具体的な利用方法を規定するものではなく、機械学習（ML）を含むAIの活用に関する指針原則を確立し、推進することを目的としています。

体系的な安全性モニタリングの開始は、インターネットや電子的報告手段が広く普及する以前にさかのぼります。そのため、当初は主として手作業に依存したプロセスでありコンピュータは主にデータ集計などの目的で使用されていました。

個別症例安全性報告（ICSR）はPVの中核的構成要素であり、市販後安全性監視の基盤であり続けています。ICSRは、承認された医薬品に関する重要な安全性情報を提供し、より広範なシグナル管理システムの中で評価されることにより、患者への危害を軽減するうえで重要な役割を果たします。

ICSR処理には、収集、トリアージ、データ入力、品質レビュー、医学的評価、さらに他の安全性データベース（例：規制当局）への報告といった複数の工程が含まれます。製品承認数および患者曝露の増加に伴い、有害事象（AE）報告数も増加します。ICSR件数の増大に加え、PVに影響を与える厳格な規制要件が存在することにより、ICSR処理およびコンプライアンス対応には重大な課題が生じています。

有害事象（AE）報告の個別分析または集積分析の結果としてシグナルが検出された場合それは、シグナルトリアージ、バリデーション、ならびに科学的評価に基づく、仮説検証研究などの独立したデータセットを用いた正式な評価を含む一連の段階を通じて、体系的に調査される必要があります。<sup>11</sup> こうした調査は、利用可能なすべての科学的エビデンスと論理を統合した全体論的な形で実施されなければならない、データインサイトのためにAIを活用するより広い機会を提供します（[図1](#)参照）。

AE報告の分析における従来のPV手法には、以下が含まれます：<sup>12,11</sup>

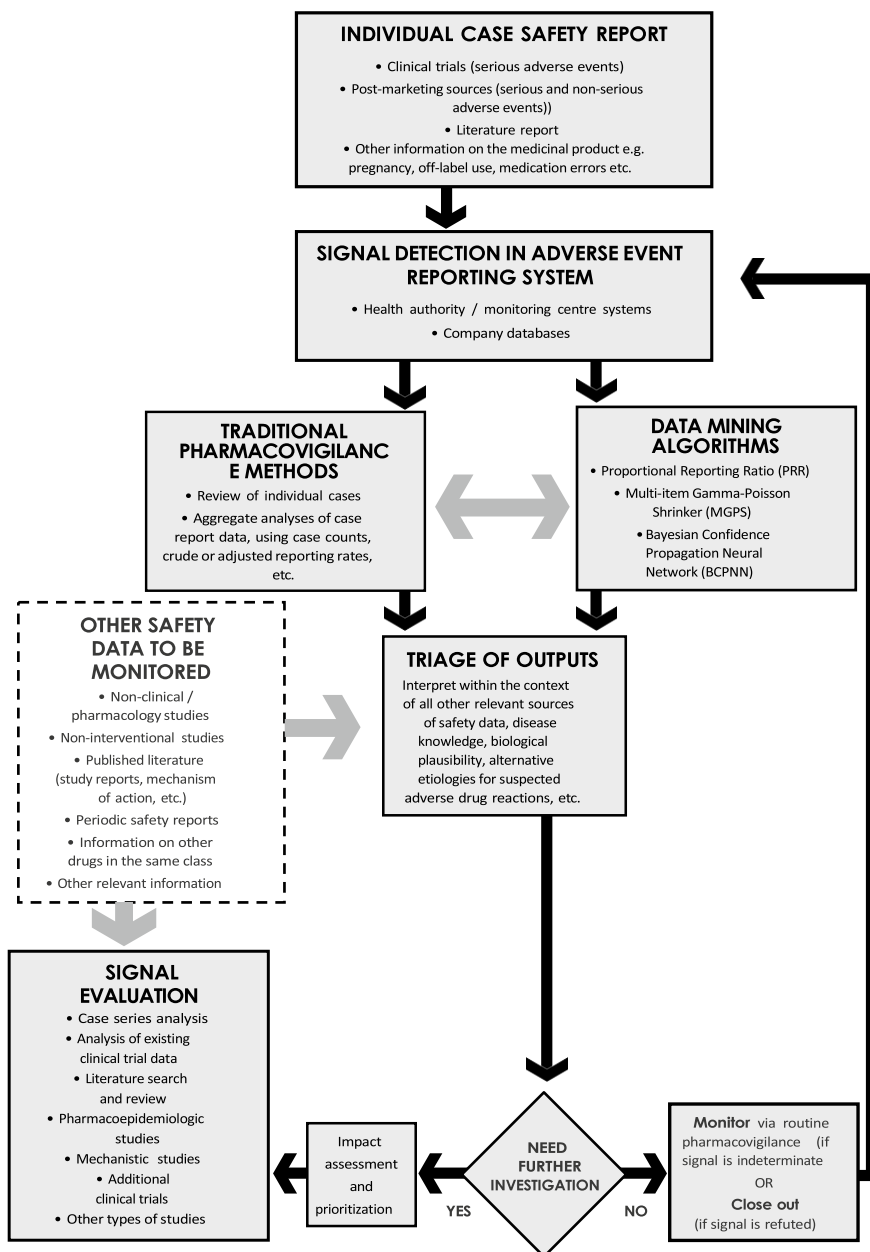
- データベース内、または公表された医学・科学文献における、ICSRまたは症例シリーズのレビュー；ならびに
- 絶対症例数、単純な報告率、割合、または推定曝露調整報告率を用いた、症例報告の集積分析。

ICSRはPVの基盤ですが、PVライフサイクル全体を通じて、他のデータストリームも考慮されます。これらのストリームは、直接リンクされている場合もあれば概念的に関連している場合もあり、薬物動態／薬力学（PK-PD）データ、リアルワールドデータ（RWD）、文献、臨床試験からの情報等が含まれます。

安全性上の懸念（重要な特定されたリスクまたは重要な潜在的リスクを含む）および不足情報が特定された場合、それらを幅広いステークホルダーへ適切に伝達するためのリスクマネジメント活動が実施されます。これは、集積報告書、リスクマネジメント計画、添付文書情報、および医療従事者向け直接情報提供（Direct Healthcare Professional Communications: DHPCs）などの文書を通じて達成されます。

図1：代表的なシグナル管理プロセス

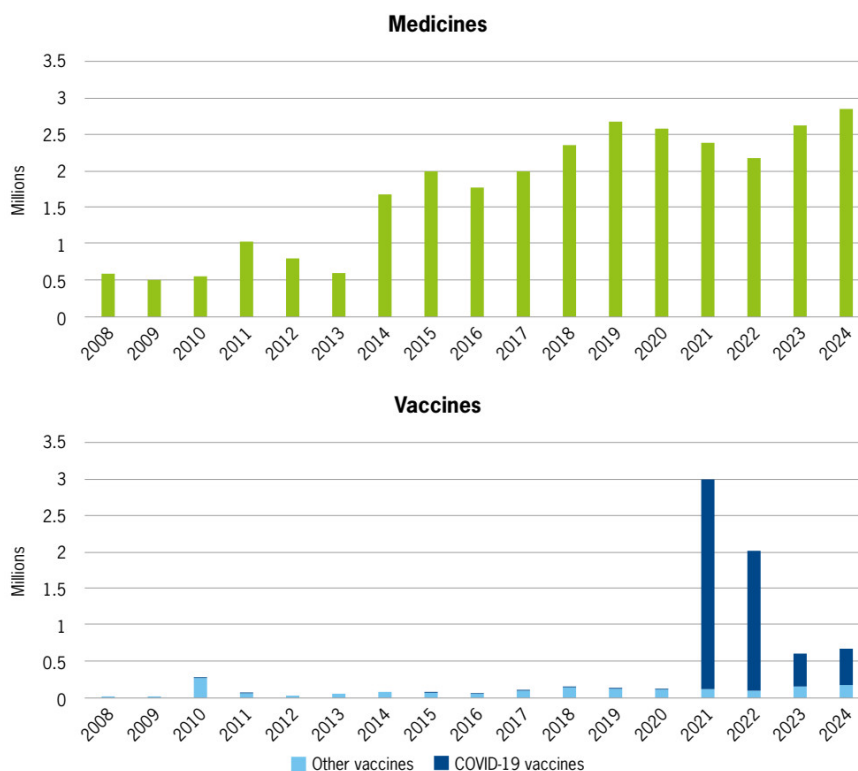
出典：Modified from CIOMS Working Group report VIII<sup>1)</sup>



COVID-19パンデミックは、PVにおける高度な手法の必要性をさらに強調しました。これは、安全性報告件数の著しい増加をもたらしたためです（図2および図3）。<sup>13,14</sup> 医薬品安全性に対する社会的認識および期待が引き続き高まる中で、医薬品に関連する潜在的风险を効果的に特定し、軽減できる頑健なPVシステムへの需要が一層高まっています。

図2：医薬品とワクチンの有害事象報告に関する世界保健機関（WHO）のグローバルデータベースであるVigiBaseの経時的成長

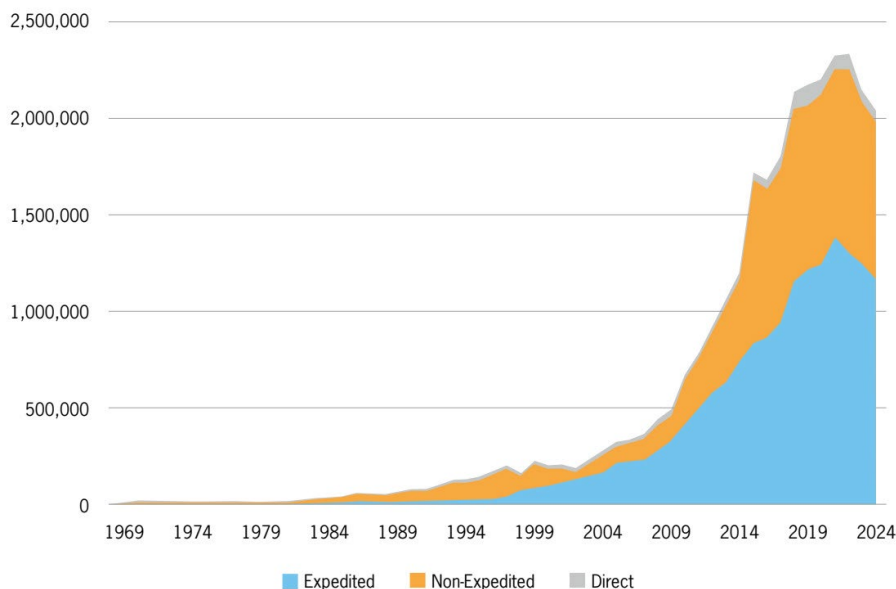
出典：VigiBase accessed April 2025. Figure reproduced with permission.



VigiBase (2025年4月アクセス) で年間に受領された医薬品およびワクチン別の報告数。

### 図3：FDA有害事象報告システム (FAERS) データベースの経時的成長

出典：Constructed using FDA FAERS database.<sup>15</sup>



データベース (2025年10月アクセス) に毎年受理された報告の数は、緊急報告、非緊急報告、FDAに直接提出された報告に分類されています。これには、例えばフォローアップ報告の形で、同じ事例の複数のバージョンが含まれる場合があります。

世界的に多様で変化し続ける規制環境の中で、ますます複雑化するPVシステムを構築し維持することの課題は増大しています。一方では症例数の増加や規制要件の複雑化といった既存の圧力があり、他方ではデータソースの増加が進んでいることから、従来のPV戦略を再考する必要があります。

技術的ソリューションは、すでにPVの進化に不可欠な要素となっています。技術が変革を可能にする基盤であるという考え方は製品開発のあらゆる領域に及びますが、革新的な自動化ツールおよびプロセスをPVへ適用することは、もはや選択肢ではなく不可欠な要件であることが明らかです。

#### 人工知能の急速な進化

従来型のAI手法 (例：K-meansクラスタリング、決定木 (decision tree)、サポートベクターマシン) は特定のタスクに合わせてカスタマイズされ、主として教師あり学習技術を活用するものでした。それに対して、BERT<sup>i</sup>に代表されるディープニューラルネットワークは自然言語処理 (NLP) において重要な役割を果たしてきました。これらは大規模なデータセットで事前学習された後、特定の用途に合わせてファインチューニングされ、予測可能な出力を提供します。

<sup>i</sup> BERT: トランスフォーマーによる双方向エンコーダ表現

しかし、生成AI（GenAI）、ナレッジグラフ、オントロジーといった新興技術の登場により、状況はこの枠組みを超えて進化しています。生成AIモデルは、広大で多様なテキストコーパス（大量の言語データ）で学習され、多くの場合、人間のフィードバックによる強化学習のフェーズが組み込まれています。これらのモデルは、高度なプロンプトを使用することで、ゼロショット（事前の事例提示なし）またはフューショット（わずかな事例提示）の機械学習（ML）技術を採用し、特定のタスクを実行することができます。

## 1.1. 適用範囲

本文書は、PVに従事する者に加え、規制当局、医薬品産業の専門家、ソフトウェアベンダー、国際および国内のPV組織、研究者、医療従事者など、PV領域向けAIソリューションを開発する組織を対象として指針を提供することを目的としています。

本報告書は、技術的な詳細ガイダンスを提示するものではなく、PVにおけるAIの統合および実装のための原則およびベストプラクティスから成る包括的な枠組みを提案します。AI技術の急速な進化と応用拡大を踏まえ、CIOMS Working Group XIVは、PV活動へのAIシステムの開発および統合を導くことを目的として本文書を作成しました。

本文書の対象範囲は、あらゆる情報源から得られる個別症例安全性報告（ICSR）の最適な収集、整理、分析および伝達に関する、直接的および間接的なすべての側面に焦点を当てています。これには、リアルワールドデータ（RWD）、医学文献、無作為化比較試験（RCT）ソーシャルメディアが含まれます。さらに、安全性データベースの検索性を向上させるシステムや、安全性データベースへのより迅速で効果的かつ一貫したデータ入力を可能にする機能など、PVに密接に関連する生産性向上ツールも含まれ、<sup>16</sup> これらは安全性監視の向上にも寄与します。<sup>17</sup>

一方で、本文書の対象範囲には、PVの直接的な管轄を超える広範な医療データ活用、例えば薬剤疫学や、ICSRの領域外に位置付けられるその他のリアルワールドエビデンス研究の設計および実施は含まれていません。同様に、PV活動と直接関連しないAIの一般的な生産性向上ツールとしての使用（例：電子メール支援など）も、考慮すべき事項が異なる可能性があるため対象外としています。

本文書の対象範囲は、本文書の寿命を長く保つために、詳細なガイダンスではなく、AIPVにおけるAIの原則とその応用として整理された実用的な指針を提供することに、意図的に限定されています。AIは極めて速い速度で進化しているため、将来展望および考慮すべき事項については後続の章で説明します。

## 第1章 – 参考文献

- 1 Organisation for Economic Co-operation and Development (OECD). Explanatory memorandum on the updated OECD definition of an AI system. OECD Artificial Intelligence Papers, No. 8. Paris: OECD Publishing; 2024. <https://doi.org/10.1787/623da898-en>.
- 2 Hauben M. Artificial Intelligence in pharmacovigilance: Do we need explainability? *Pharmacopidemiol Drug Saf.* 2022;Dec;31(12):1311-1316. <https://doi.org/10.1002/pds.5501> (Journal full text)
- 3 Kompa B, Hakim JB, Palepu A, Kompa KG, Smith M, Bain PA, Woloszynek S, Painter JL, Bate A, Beam AL. Artificial intelligence based on machine learning in pharmacovigilance: A scoping review. *Drug Saf.* 2023;Apr46(4):433. <https://doi.org/10.1007/s40264-023-01273-9> (Journal full text)
- 4 Webb S. Deep learning for biology. *Nature.* 2018;Feb22;554:555-557. <https://doi.org/10.1038/d41586-018-02174-z> (Journal abstract)
- 5 Noorbakhsh-Sabet N, Zand R, Zhang Y, Abedi V. Artificial intelligence transforms the future of health care. *Am J Med.* 2019;Jul1;132(7):795-801. <https://doi.org/10.1016/j.amjmed.2019.01.017> (Journal full text)

- 6 Wright T, Yan P, Easterbrook M. Machine learning to identify multifocal ERG deficits in patients taking hydroxychloroquine. *Invest Ophthalmol Vis Sci.* 2019;Jul22;60(9):5959. ([Journal abstract](#) accessed 15 October 2025)
- 7 Chang DW, Lin CS, Tsao TP, et al. Detecting digoxin toxicity by artificial intelligence-assisted electrocardiography. *Int J Environ Res Public Health.* 2021;Apr6;18(7). <https://doi.org/10.3390/ijerph18073839> ([Journal full text](#))
- 8 Li MH, Mestre TA, Fox SH, Taati B. Vision-based assessment of parkinsonism and levodopa-induced dyskinesia with pose estimation. *J Neuroeng Rehabil.* 2018;Dec15;15(1):3. <https://doi.org/10.1186/s12984-018-0446-z> ([Journal abstract](#))
- 9 Pathak A, Serrer L, Zapata D, King R, Mirel LB, Sukalac T, et al. Privacy preserving record linkage for public health action: opportunities and challenges. *J Am Med Inform Assoc.* 2024;Nov1;31(11):2605-2612. <https://doi.org/10.1093/jamia/ocae196> ([Journal abstract](#))
- 10 Eisinger-Mathason TSK, Leshin J, Lahoti V, Benner B, Snyder M, Kohane I, et al. Data linkage multiplies research insights across diverse healthcare sectors. *Commun Med.* 2025;5:58. <https://doi.org/10.1038/s43856-025-00769-y> ([Journal full text](#))
- 11 Council for International Organizations of Medical Sciences (CIOMS). Practical aspects of signal detection in pharmacovigilance: CIOMS Working Group VIII report. Geneva: Council for International Organizations of Medical Sciences; 2010. ([Full text](#) accessed 15 October 2025)
- 12 U.S. Food and Drug Administration (FDA). Good pharmacovigilance practices and pharmacoepidemiologic assessment: guidance for industry. Silver Spring (MD): U.S. Food and Drug Administration; 2005;Mar. ([Full text](#) accessed 21 March 2025).
- 13 Alvager T, Smith TJ, Vijai F. Neural-network applications for analysis of adverse drug reactions. *Biomed Instrum Technol.* 1993;Sep-Oct;27(5):408-411. ([Journal abstract](#) accessed 15 October 2025)
- 14 Kěpuska V, Bohouta G. Next-generation of virtual personal assistants (Microsoft Cortana, Apple Siri, Amazon Alexa and Google Home). In: 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC). Las Vegas (NV): IEEE; 2018. p. 99-103. <https://doi:10.1109/CCWC.2018.8301638>. ([Journal abstract](#))
- 15 U.S. Food and Drug Administration (FDA). FDA Adverse Event Reporting System (FAERS) Public Dashboard.[Internet]. Silver Spring (MD): U.S. Food and Drug Administration; 2025. ([Webpage](#) accessed 29 October 2025)
- 16 Painter JL, Chalamalasetti VR, Kassekert R, Bate A. Automating pharmacovigilance evidence generation: using large language models to produce context-aware structured query language. *JAMIA Open.* 2025;Feb;8(1):oaf003. <https://doi.org/10.1093/jamiaopen/oaf003> ([Journal full text](#))
- 17 Painter JL, Mahaux O, Vanini M, Kara V, Roshan C, Karwowski M, Chalamalasetti VR, Bate A. Enhancing drug safety documentation search capabilities with large language models: a user-centric approach. In: Proceedings of the 2023 International Conference on Computational Science and Computational Intelligence (CSCI). Las Vegas (NV): IEEE; 2023. p. 1005-1010. 2023;Dec13:49-56. <https://doi:10.1109/CSCI62032.2023.00015> ([Journal abstract](#))

## 第2章 ランドスケープ分析

### 2.1. これまでの医薬品安全性監視における人工知能の利用

人工知能（AI）は、医薬品安全性監視（PV）のあらゆる側面に直接的または間接的な影響を及ぼし得ます（図1：代表的なシグナル管理プロセス）。本章では、AIの要素を取り入れ、PV全体にわたるさまざまな業務のために開発または導入されてきたシステムについて説明します。特に、PV向けに具体的に実装されたもの、またはPVのアプリケーションやプロセスにおいて顕著な特性や機能を有するものに焦点を当てます。

例えば、一般的な翻訳業務を目的としたAIシステムは本書の対象範囲外ですが、有害事象（AE）報告や報告者の原資料（例：原症例報告書、通話記録の書き起こし）など、PVに特化した翻訳は対象範囲に含まれます。また、疫学研究における傾向スコアモデルへ組み込む共変量を特定するためのAI手法に関する研究も、本書の対象外としています。

本章の目的は、網羅的な一覧を提示することではなく、現在のAI応用の範囲と多様性を示すことにあります。追加の事例については最近のレビュー論文を参照してください。<sup>1</sup> また、読者はPVにおけるAI利用を論じた多くの展望論文や解説、<sup>2,3,4,5</sup> および提示されている注意喚起についてもあわせて参照することが推奨されます。<sup>6</sup>

#### 2.1.1. 有害事象の捕捉

医薬品使用に関する（個人的経験としての）言及や、有害事象（AE）への言及を特定し、それらがAE報告の基礎となり得る情報を抽出することを目的として、ソーシャルメディアコンテンツに対する自然言語処理（NLP）に関連するさまざまなタスクに対し、AIシステムが提案および評価されています。これらのタスクには、関連投稿の特定、<sup>7,8</sup> 投稿中の関連部分の抽出、<sup>9</sup> 投稿内で記述されたAEまたは医薬品の内容を、MedDRA（Medical Dictionary for Regulatory Activities）やATC（Anatomical Therapeutic Chemical）分類コードなどの標準化用語へ正規化する処理、<sup>10</sup> さらに同一投稿内で言及されたAEと医薬品との関係を分類する処理<sup>11</sup> が含まれます。

同様に、AE報告に取り込まれる可能性のある有害事象を特定するため、科学文献のスクリーニングを支援するAI手法も開発されています。<sup>12,13</sup>

#### 2.1.2. 個別症例安全性報告処理

ICSR処理の分野において、AIシステムが少なくとも2010年代以降、一部の組織で日常的に使用されてきた領域の一つが重複検出です。これは、特定の患者における同一の有害事象（AE）を記述しているにもかかわらず、互いに関連付けられていない複数の記録を識別することに関係します。<sup>14</sup> 機械学習（ML）および確率的レコードリンケージに基づく重複検出手法は、VigiBase、<sup>15</sup> 米国食品医薬品局（FDA）の有害事象報告システム（FDA Adverse Event Reporting System: FAERS）、<sup>16</sup> およびEudraVigilance<sup>17</sup> において実装されています。さらに、自由記述テキストから情報を抽出して取り込むことにより重複

検出を改善する目的で、NLPの活用も検討されています。<sup>16</sup> ルールベース手法はより広く使用されており実装も容易ですが、性能面ではこれらの手法に劣ることが示されています。<sup>14,18</sup>

ICSR処理を支援する目的でAIが活用されているもう一つの領域は、有害事象<sup>19,20</sup> または医薬品<sup>21</sup> に関する情報を、逐語記載欄 (verbatim fields) や自由記述形式の症例ナラティブに基づいて、標準用語体系へコード化する作業です。NLPはまた、症例ナラティブから関連情報を抽出し、それを構造化フィールドへマッピングする用途<sup>22,23,24,25,26,27</sup> や、ICSR翻訳にも適用されています。<sup>28</sup>

多数の症例報告を処理するいくつかの組織では、ルールベースのいわゆるロボティック・プロセス・オートメーション (RPA) 技術を用いて、反復的かつ労働集約的な作業を自動化しています。<sup>29</sup> これらの技術は、他のコンピュータシステムのユーザーインターフェイス上で動作し、人間が通常実施する操作を模倣します。<sup>30</sup> 例えば、TransCelerate が示したように、重複確認や症例の取り込み作業の自動化などが挙げられます。<sup>31</sup>

ICSR処理におけるAIシステムのその他の応用としては、構造化データから症例ナラティブを生成する手法、<sup>29</sup> 受信報告のヒトレビュー優先度付け (トリアージ) 支援、<sup>32,33</sup> 個別症例の因果関係評価支援、<sup>34</sup> ならびに症例ナラティブ中の個人名を自動的にマスキング (匿名化) する技術などが開発されています。<sup>35</sup>

### 2.1.3. シグナル検出および分析

PVにおけるAIの実環境での利用例として最も初期のものは、1990年代後半にさかのぼります。当時、1970年代に概念化された不均衡分析 (disproportionality analysis) が、<sup>36</sup> ルールベースのトリアージアルゴリズムの一部として導入され、大規模な国内および国際的な個別症例報告の解析において、PV専門家の注意を向ける支援として活用され始めました。<sup>37,38,39,40</sup> その後、回帰分析や<sup>41,42</sup> 傾向スコア (propensity score) <sup>43</sup> による交絡調整、自薬剤間相互作用への拡張、<sup>44,45,46,47,48,49</sup> ならびに有害反応に関与し得るその他のリスク因子の評価など、さまざまな段階的改良が導入および評価されてきました。<sup>50</sup> また、製造工程に関連する有害事象や、品質不良または偽造医薬品に関連する有害事象を検出する手法も検討されています。<sup>51,52,53</sup> さらに、症例シリーズ (case series) の地理的分布、個別報告の内容および品質、<sup>54</sup> 発現までの時間 (time-to-onset) 、<sup>55</sup>あるいはNaranjoスコアと、医療従事者 (HCP) および販売承認保有者 (MAH) から報告された割合の組み合わせなどを考慮した統計的シグナル検出のための予測モデル開発も試みられています。<sup>56</sup>

NLPは、既知または未知で潜在的に重篤な有害作用に関する情報を抽出する目的で、規制情報、<sup>57</sup> 科学文献、ならびに臨床記録<sup>58,59,60,61</sup> の解析に適用されています。これにより、特に初期シグナル評価および優先順位付けの段階における意思決定の支援および効率化が期待されています。

AIに基づくシグナル検出の研究の中には、真に新規の有害事象を発見し得る可能性を示唆する興味深い結果も報告されています。<sup>62</sup> 一方で、高度かつ洗練された手法は、結果の過度に楽観的な解釈や過剰なメッセージングにつながる可能性があり、それらが広く拡散するリスクがあるため、慎重な評価が求められます。<sup>63</sup>

いくつかの組織では、薬剤と有害事象との因果関連の評価や規制対応判断の支援を目的として、ICSR優先順位付けのための予測モデルが開発されています。<sup>64,65,33,66,67</sup> これらはシグナル評価または症例処理において人間によるレビュー対象を優先的に選定するために利用されます。また、症例ナラティブに対するセマンティック検索 (semantic search) は、シグナル検出および評価の支援を目的として開発されており、<sup>68,69</sup> さらに、シグナル検証を支援するMLベースの意思決定支援<sup>70</sup> や、症例報告中の関連情報を自動的に可視

化して人間のレビューを補助する取り組みも行われています。<sup>71</sup> MLはまた、ICSRに報告された表現型情報に基づき、薬剤毒性に関連する遺伝子型を有する患者割合の推定にも活用されています。<sup>72</sup>

教師なし学習 (unsupervised learning) の応用も、特に類似または関連する有害事象を記述する報告を統合的に解析する目的で開発されています。これには、有害事象 (および限定的ではあるが薬剤) のネットワーク解析<sup>73,74,75,76,77</sup>、有害事象報告のクラスタ解析<sup>78,79</sup>、ならびに有害事象および薬剤のセマンティック表現 (意味表現) をデータ駆動で導出する手法が含まれます。<sup>80,81</sup>

DrugBank<sup>82</sup> やSIDER<sup>83</sup> のような副作用および適応症情報を含むデータセット、さらにBio2RDF<sup>84</sup> のような薬理学的情報および化学構造データを含むデータセットも、PVにおけるシグナル検出および分析の高度化に活用されています。<sup>85,86,87</sup> これらはAIベースの予測的シグナル検出の下流インプット (下流工程の入力データ) として利用可能な知識グラフ (Knowledge graphs) の構築にも寄与しています。<sup>88,90</sup> また、潜在的な有害作用の分析に関連する科学論文の検索を支援するAIの応用も開発されています。

#### 2.1.4. 医薬品安全性監視における生成AIの初期応用

医薬品安全性監視 (PV) における生成型大規模言語モデル (large language models: LLMs) の初期的応用については、すでに検討が開始されています。これには、生成型LLMの出力をプロンプト設計または後処理によって制限し、分類や固有表現認識 (named entity recognition) などの基盤として利用する応用も含まれます。

これまでに報告されている例としては、規制当局から患者へのコミュニケーション文書を平易化する用途、<sup>89</sup> 医薬品ラベリング文書の要約、<sup>90</sup> 症例ナラティブの要約、<sup>91</sup> 科学文献およびソーシャルメディアのスクリーニング、<sup>13,92</sup> 医薬品安全性関連文書の検索、<sup>5</sup> ラベリング情報に関するQ&A対応、<sup>93</sup> PV文脈を理解した構造化照会言語 (Structured Query Language: SQL) コード生成、<sup>94</sup> ならびに報告者へのフォローアップレター草案作成などがあります。<sup>95</sup>

#### 2.1.5. 実運用されているAIソリューションの例

これまでの医薬品安全性監視 (PV) におけるAIソリューションの研究開発の多くは実験的段階にとどまっており、実環境での導入がまだ行われていないもの、あるいはパイロットスタディなど限定的な試験的利用にとどまっているものが多い状況です。しかしながら、表1には、複数のPV組織において日常業務として採用され、かつ公開情報として記載されているAIソリューションの例が示されています。製薬企業によるAIソリューションの導入は、主としてソフトウェアベンダーによる実装に基づいている場合が多いと考えられますが、これらの詳細は公開情報としては必ずしも記載されていません。<sup>96</sup> 同様に、欧州医薬品庁 (EMA) において運用されている複数のAIソリューションについても公開情報で紹介されていますが、個別の科学論文としてはまだ公表されていないものがあります。<sup>97</sup> さらに、[付録3](#) に示されているユースケースも参照してください。

表1: パブリックドメインで記述された医薬品安全性監視における人工知能ソリューションの導入例

出典: CIOMS XIV working group

AIソリューション	PVコンテキスト／データベース
医薬品の自動コーディング	VigiBase <sup>21</sup>
重複検出	FAERS <sup>98</sup> 、VigiBase <sup>15</sup>
個別症例報告の自動トリアージ	Swedish Medical Products Agency <sup>34</sup> 、製薬企業 <sup>99</sup>
定量的シグナル検出のための自動トリアージ	様々な規制当局、国際機関、製薬企業のデータベース
定量的シグナル検出のための予測モデル	VigiBase <sup>56,100</sup> 、オランダ医薬品安全性監視センターLareb <sup>101</sup>
シグナル検出と評価のための有害事象のクラスター解析	VigiBase <sup>80,102</sup>
安全性データに関する文献調査	EudraVigilance <sup>103</sup> 、オランダ医薬品安全性監視センターLareb <sup>104</sup>

## 2.2. 規制上の考慮事項

2017年以降、世界各国では、技術進歩およびそれが社会や経済に与える影響に対応するため、国家レベルのAI戦略の策定が進められてきました。<sup>105</sup> 各国は、AIシステムの倫理的利用および信頼性を確保することを目的として、それぞれ異なる規制枠組みおよび指針原則を整備しており、AIに関する法制度の導入も進められています（例：欧州人工知能法 [EU AI Act]<sup>106</sup>、カナダ人工知能・データ法 [Artificial Intelligence and Data Act: AIDA]<sup>107,111</sup>）。さらに、欧州医薬品庁（EMA）および米国食品医薬品局（FDA）により、医薬品分野におけるAI利用に関するリフレクションペーパーおよびディスカッションペーパーが公表されています。また、米国FDAからは、医薬品の規制上の意思決定を支援するAI利用に関するドラフトガイダンスも公開されています。

### 2.2.1. 医薬品安全性監視におけるAIの指針原則

政府、規制当局、ならびに世界保健機関（WHO）や経済協力開発機構（OECD）などの国際機関により、安全かつ責任あるAI利用のための指針原則（guiding principles）が数多く公表されています。CIOMS Working Group XIV は、規制領域における安全で責任あるAI利用のための指針原則および推奨ベストプラクティスを定義した代表的な出版物をレビューしました。これらの文書はPVを対象として作成されたものではありませんが、記載されているAI指針原則はPV分野にも適用可能であると判断されました。なお、これら複数の出版物では、一部の原則が他の原則と組み合わせられて記述されている場合があるため、PVに適用する指針原則を整理・確立するにあたり、CIOMS Working Group XIV による一定の裁量が用いられています。特筆すべき点として、CIOMS Working Group XIV が参照した米国の文書の一部は現在アーカイブ化されており、最新の米国政策を必ずしも反映していません。しかしながら、本報告書における指針原則の策定に使用された経緯を踏まえ、当該参考文献は保持されています。各国における最新のAI政策については、Artificial Intelligence Policy Trackerを参照してください。<sup>108</sup> 表2では、主要な政府機関および国際機関におけるAI指針原則の全体的比較を示しています。

AI原則の概要説明（網羅的ではありません）は、[付録2](#)に記載されています。

**表2：地域および国の政府機関、国際機関におけるAI指針原則の比較**  
－ CIOMS Working Group XIV作成

出典：CIOMS Working Group XIV

原則	地域・国の政府機関や国際機関の原則の例								
	EU <sup>109, 110</sup>	オーストラリア <sup>111</sup>	カナダ <sup>113</sup>	シンガポール <sup>112</sup>	UK <sup>113</sup>	US <sup>114</sup>	PAHO <sup>115</sup>	WHO <sup>116</sup>	OECD <sup>117</sup>
ヒューマン・オーバーサイト	✓	✓	✓		✓	✓	✓	✓	✓
妥当性と頑健性	✓	✓	✓		✓		✓	✓	✓
データプライバシー	✓	✓			✓	✓	✓		
透明度	✓	✓	✓	✓	✓		✓	✓	✓
説明責任	✓	✓	✓	✓	✓	✓	✓	✓	✓
社会的ウェルビーイング	✓	✓		✓			✓	✓	✓
環境的ウェルビーイング	✓	✓						✓	✓
公正性と衡平性	✓	✓	✓	✓	✓	✓	✓	✓	✓
説明可能性	✓	✓		✓	✓	✓		✓	✓
安全性	✓	✓	✓		✓	✓		✓	✓
ガバナンス	✓				✓			✓	

## 2.2.2. 医薬品ライフサイクルにおける人工知能（AI）利用に関するEMAリフレクションペーパー

2024年9月9日、欧州医薬品庁（EMA）は「医薬品ライフサイクルにおけるAI利用に関するリフレクションペーパー」を最終化しました。<sup>118</sup> 本リフレクションペーパーは、医薬品の安全かつ有効な開発、製造、および使用におけるAI／機械学習（ML）の利用を対象としています。

EMAは、システムライフサイクル全体を通じたAIおよびMLツールの開発、導入、および監視において、リスクに基づくアプローチを採用することを提唱しています。本文書では、患者安全に影響を及ぼすシステムを「高い患者リスク（high patient risk）」、規制上の意思決定に重大な影響を与える場合を「高い規制影響（high regulatory impact）」として定義しています。申請者／販売承認保有者（MAH）およびAI・MLシステム開発者は、規制影響およびリスク分析を実施することが期待されています。AIおよびMLシステムに対する審査の厳格さは、リスク水準および規制影響の評価結果に応じて決定されます。

本文書は、創薬研究および開発から承認後段階に至るまで、医薬品ライフサイクル全体におけるAIおよびML利用に関する技術的および規制上の考慮事項を提示しています。

特にPVに関しては、AI/MLツールが、適用される製造販売後安全管理の基準（Good Pharmacovigilance Practices：GVP）要件に沿って、有害事象（AE）報告管理やシグナル検出などの活動を効果的に支援し得ることが示されています。PV領域における応用では、例えばAE報告の重篤度評価やシグナル検出の改善を目的として、他領域よりも柔軟なAI/MLモデリングおよび導入アプローチが可能となる場合があります。しかしながら、使用されるすべてのアルゴリズムおよびモデルに関連するリスクを軽減するため、モデル性能のバリデーション、監視および文書化を行い、AI/ML運用をPVシステムに組み込む責任はMAHにあります。

一般に、申請者またはMAHは、AIおよびMLアプリケーションを構成するすべての要素（すなわち、アルゴリズム、モデル、データセット、およびデータ処理パイプライン）が目的適合性を満たし、GxP基準およびEMAの最新科学的ガイドラインに適合していることを保証する責任を負います。AIシステムにおけるデータ保護コンプライアンス遵守の監督およびモニタリングは、加盟国のデータ保護当局が担います。また、特に規制上の意思決定に高い影響を及ぼす可能性のあるAI/MLモデルについては、申請者、MAH、および開発者が実験的技術段階からEMAと早期に協議することが推奨されています。<sup>118</sup>

EMAは、PVを含む医薬品ライフサイクルにおけるAI利用に関して、今後さらに詳細なガイダンスを策定する計画です。<sup>118</sup>

### 2.2.3. 医薬品および生物学的製剤開発における人工知能および機械学習の利用に関する米国FDAディスカッションペーパー

2023年5月、米国食品医薬品局（FDA）は、「医薬品および生物学的製剤開発における人工知能および機械学習の利用（Using Artificial Intelligence & Machine Learning in the Development of Drug & Biological Products）」に関するディスカッションペーパーを公表しました。<sup>119</sup> 米国FDAは、医薬品開発ライフサイクル全体においてAI/MLの利用が増加していることを認識しており、データマイニング、大規模マルチオミクス解析、PK/PDモデリング、リアルワールドデータ、ウェアラブルデバイスからのデータ収集、ならびにその他のデータセット（例：in vitro試験、in vivo試験、機序研究(mechanistic studies)、多臓器チップシステム）を活用した新たなアプローチが進展しているとしています。市販後安全性監視において、米国FDAはAI/MLが以下の用途に活用される可能性があるとして示しています。すなわち、i) 報告件数の増加およびデータソースの複雑化に対応するため、AI/MLを用いたICSR処理および優先順位付けの自動化、ii) 医薬品と有害事象（AE）との因果関係の可能性に基づくICSR分類、iii) ICSRにおける転帰の重篤性判定、iv) 特定製品に関連する複数有害事象についての集積報告（aggregate reports）の自動作成です。

### 2.2.4. 医薬品および生物学的製剤における規制意思決定を支援する人工知能利用に関する米国FDAドラフトガイダンス

米国FDAは2025年1月、「医薬品および生物学的製剤における規制意思決定を支援する人工知能利用に関する考慮事項（Considerations for the Use of Artificial Intelligence to Support Regulatory Decision-Making for Drug and Biological Products Guidance for Industry and Other Interested Parties）」と題するドラフトガイダンスを公表しました。<sup>120</sup>。本ガイダンスでは、AIに対するリスクベースの信頼性評価（credibility assessment）フレームワークが詳述されています。本ガイダンスの対象範囲は、医薬品の安全性、有効性、または品質に関する規制意思決定を支援する用途に限定されています。一方で、創薬段階や、業務効率化のみを目的としてAIを導入するケースは対象外です。また、本ガイダンスではAIを広義に捉えており、機械学習（ML）など特定のAI技術に限定していません。本ドラフトガイダンスは、主として以下の三つの主要構成から成り立っています。

1. リスクに基づく信頼性評価フレームワークの確立（第3章「リスクに基づくアプローチ」参照）
2. ライフサイクルを通じた信頼性維持
3. AIモデル開発について当局と協議するためのスポンサー向け選択肢

リスクベースド信頼性評価フレームワークは、以下の7つのステップで構成されています。

1. AIモデルによって解決すべき問いを定義する。
2. 「使用文脈（Context of Use：COU）」を定義する。COUとは、「関心のある問いに対応するために使用されるAIモデルの具体的な役割および適用範囲」を指します。重要な点として、ここには回答される質問、およびその後生じる分類や決定がAI出力のみに基づくのか、それとも他の情報と併用されるのか（すなわち“モデル影響度”）が含まれます。これは次段階におけるリスク定義に重要な要素となります。
3. モデルリスクを定義する。これはCOUで定義されたモデル影響度および意思決定結果の影響、すなわち誤った判断がもたらす結果によって決定されます。AIモデルがスタンドアロンで稼働し、誤判断が重大な危害をもたらす場合、リスクは最も高くなります。開発および運用全体を通じた監督水準は、このリスクに比例して高まります。
4. COUにおけるAIモデルの信頼性確立計画を策定する。
5. 当該計画を実行する。
6. 信頼性評価計画の結果を文書化し、計画からの逸脱があれば説明する。
7. COUに対するAIモデルの適切性（Adequacy）を判断する。<sup>121</sup>

## 2.2.5. 米国 FDA Emerging Drug Safety Technology Program

米国FDAは2024年6月、医薬品安全性監視（PV）および医薬品ライフサイクルにおいて使用されるAIおよびその他の新興技術について産業界の関係者と連携することを目的として、Emerging Drug Safety Technology Program（EDSTP）を設立しました。EDSTPの三つの目的には、産業界と米国FDAとの間の議論の実施、新興AI／機械学習（ML）モデルおよびその他の新規技術に関する知識の普及、ならびにPVの将来的な規制または政策開発に資する情報提供が含まれます。<sup>122</sup>

## 2.2.6. 大規模言語モデルの利用に関するガイダンス

2022年11月30日にChatGPT（Generative Pre-trained Transformer）が公開されて以降、生成AI（GenAI）をさまざまな業務（文章および画像生成、コーディング、ブレインストーミング、調査など）へ適用し、生産性向上を図る可能性について多くの検討が進められてきました。GenAIの潜在的利用可能性および広範な適用性を踏まえ、規制当局および各種組織は、それぞれの職員およびより広いステークホルダーによるGenAIの安全かつ責任ある利用を目的としたハイレベルガイドおよびベストプラクティスを策定しています。これらは、既存のAIに関する指針原則と整合しています。

- 医薬品規制科学および医薬品規制業務における大規模言語モデル（LLM）の利用に関する指針原則（EMA<sup>123</sup>）
- 生成AIの利用に関するガイド（カナダ<sup>124</sup>）
- 生成的人工知能に関する初期政策的考慮事項（OECD<sup>125</sup>）
- WHO「健康のための人工知能の倫理およびガバナンス：大規模マルチモーダルモデルに関するガイダンス（WHO Ethics and governance of artificial intelligence for health: Guidance on large multi-modal models（WHO<sup>126</sup>）

## 2.2.7. 安全な人工知能に関するガイドライン

安全なAIに関しては、WHOおよびOECDなどの規制当局および国際機関から関連するガイドラインが公表されています。

- 医療分野における人工知能に関する規制上の考慮事項（WHO, 2023）<sup>127</sup>
- 信頼できるAIのための倫理ガイドライン（欧州委員会, 2019）<sup>128</sup>
- 人工知能に関する理事会勧告（OECD, 2019、2023年改訂）<sup>129</sup>
- 医療機器開発における Good Machine Learning Practice：指針原則（米国FDA、Health Canada、MHRA, 2021）<sup>130</sup>
- 機械学習対応医療機器における透明性：指針原則<sup>131</sup>
- ISO/IEC 23894:2023 情報技術—人工知能—リスクマネジメントに関するガイダンス<sup>132</sup>

## 第2章 – 参考文献

- 1 Salas M, Petracek J, Yalamanchili P, Aimer O, Kasthuril D, Dhingra S, Junaid T, Bostic T. The use of artificial intelligence in pharmacovigilance: a systematic review of the literature. *Pharm Med.* 2022;Oct;36(5):295-306. <https://doi.org/10.1007/s40290-022-00441-z> (Journal abstract)
- 2 Basile AO, Yahi A, Tatonetti NP. Artificial intelligence for drug toxicity and safety. *Trends Pharmacol Sci.* 2019;Sep1;40(9):624-635. <https://DOI:10.1016/j.tips.2019.07.005> (Journal full text)
- 3 Kassekert R, Grabowski N, Lorenz D, Schaffer C, Kempf D, Roy P, Kjoersvik O, Saldana G, ElShal S. Industry perspective on artificial intelligence/machine learning in pharmacovigilance. *Drug Saf.* 2022;May;45(5):439-448. <https://doi:10.1007/s40264-022-01164-5> (Journal full text)
- 4 Bate A, Stegmann JU. Artificial intelligence and pharmacovigilance: What is happening, what could happen and what should happen?. *Health Policy and Technology.* 2023;Jun1;12(2):100743. <https://doi.org/10.1016/j.hlpt.2023.100743> (Journal full text)
- 5 Painter JL, Kassekert R, Bate A. Corrigendum: An industry perspective on the use of machine learning in drug and vaccine safety. *Front Drug Saf Regul.* 2023;Dec4;3:1244115. <https://doi.org/10.3389/fdsfr.2023.1110498> (Journal full text)
- 6 Hauben M, Reynolds R, Caubel P. Deconstructing the Pharmacovigilance Hype Cycle. *Clin Ther.* 2018; Dec;40(12):1981-1990.e3. <https://doi:10.1016/j.clinthera.2018.10.021> (Journal full text)
- 7 Freifeld CC, Brownstein JS, Menone CM, Bao W, Filice R, Kass-Hout T, Dasgupta N. Digital drug safety surveillance: monitoring pharmaceutical products in twitter. *Drug Saf.* 2014;May;37:343-350. <https://doi.org/10.1007/s40264-014-0155-x> (Journal full text)
- 8 Alvaro N, Conway M, Doan S, Lofi C, Overington J, Collier N. Crowdsourcing Twitter annotations to identify first-hand experiences of prescription drug use. *J Biomed Inform.* 2015;Dec1;58:280-287. <https://doi.org/10.1016/j.jbi.2015.11.004> (Journal full text)
- 9 Nikfarjam A, Sarker A, O'connor K, Ginn R, Gonzalez G. Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. *J Am Med Inform Assoc.* 2015;May1;22(3):671-681 <https://doi.org/10.1093/jamia/ocu041> (Journal full text)

- 10 Karimi S, Metke-Jimenez A, Nguyen A. CADEminer: a system for mining consumer reports on adverse drug side effects. In: Proceedings of the eighth workshop on exploiting semantic annotations in information retrieval. 2015; Oct 22; 47-50. <https://doi.org/10.1145/2810133.2810143> (Journal abstract)
- 11 White RW, Harpaz R, Shah NH, DuMouchel W, Horvitz E. Toward enhanced pharmacovigilance using patient-generated data on the internet. *Clin Pharmacol Ther*. 2014; Aug; 96(2):239-246. <https://doi.org/10.1038/clpt.2014.77> (Journal full text)
- 12 Park J, Djelassi M, Chima D, Hernandez R, Poroshin V, Iliescu AM, Domalik D, Southall N. Validation of a natural language machine learning model for safety literature surveillance. *Drug Saf*. 2024; Jan; 47(1):71-80. <https://doi.org/10.1007/s40264-023-01367-4> (Journal abstract)
- 13 European Medicines Agency (EMA). AI Observatory 2024. Amsterdam: European Medicines Agency; 2025; May 6. EMA/154528/2025. (Webpage accessed 23 September 2025)
- 14 Tregunno PM, Fink DB, Fernandez-Fernandez C, Lázaro-Bengoa E, Norén GN. Performance of probabilistic method to detect duplicate individual case safety reports. *Drug Saf*. 2014; Apr; 37:249-258. <https://doi.org/10.1007/s40264-014-0146-y> (Journal abstract)
- 15 Norén GN, Orre R, Bate A, Edwards IR. Duplicate detection in adverse drug reaction surveillance. *Data Min Knowl Discov*. 2007; Jun; 14:305-328. <https://doi.org/10.1007/s10618-006-0052-8> (Journal abstract)
- 16 Kreimeyer K, Menschik D, Winiecki S, Paul W, Barash F, Woo EJ, Alimchandani M, Arya D, Zinderman C, Forshee R, Botsis T. Using probabilistic record linkage of structured and unstructured data to identify duplicate cases in spontaneous adverse event reporting systems. *Drug Saf*. 2017; Jul; 40:571-582. DOI:10.1007/s40264-017-0523-4 (Journal full text)
- 17 Personal communication from Tom Paternoster-Howe (March 2025).
- 18 Kiguba R, Isabirye G, Mayengo J, et al. Navigating duplication in pharmacovigilance databases: a scoping review. *BMJ Open* 2024; 14:e081990. <https://doi.org/10.1136/bmjopen-2023-081990> (Journal full text)
- 19 Combi C, Zorzi M, Pozzani G, Moretti U, Arzenton E. From narrative descriptions to MedDRA: automagically encoding adverse drug reactions. *J Biomed Inform*. 2018; Aug 1; 84:184-199. <https://doi.org/10.1016/j.jbi.2018.07.001> (Journal full text)
- 20 Gurulingappa H, Mateen-Rajpu A, Toldo L. Extraction of potential adverse drug events from medical case reports. *Journal of Biomedical Semantics*. 2012; Dec; 3:1-10. <https://doi.org/10.1186/2041-1480-3-15> (Journal full text)
- 21 Meldau EL, Bista S, Rofors E, Gattepaille LM. Automated Drug Coding Using Artificial Intelligence: An Evaluation of WHODrug Koda on Adverse Event Reports. *Drug Saf*. 2022; May; 45(5):549-561. <https://doi.org/10.1007/s40264-022-01162-7> (Journal full text)
- 22 Abatamarco D, Perera S, Bao SH, Desai S, Assuncao B, Tetarenko N. Training augmented intelligent capabilities for pharmacovigilance: applying deep-learning approaches to individual case safety report processing. *Pharm Med*. 2018; 32(6):391-401. <https://doi.org/10.1007/s40290-018-0251-9> (Journal full text)
- 23 Wang W, Kreimeyer K, Woo EJ, Ball R, Foster M, Pandey A, Scott J, Botsis T. A new algorithmic approach for the extraction of temporal associations from clinical narratives with an application to medical product safety surveillance reports. *J Biomed Semantics*. 2016; Aug 1; 62:78-89. <https://doi.org/10.1016/j.jbi.2016.06.006> (Journal full text)
- 24 Kreimeyer K, Foster M, Pandey A, Arya N, Halford G, Jones SF, Forshee R, Walderhaug M, Botsis T. Natural language processing systems for capturing and standardizing unstructured clinical information: a systematic review. *J of Biomed Informatics*. 2017; Sep 1; 73:14-29. <https://doi.org/10.1016/j.jbi.2017.07.012> (Journal full text)
- 25 Pham P, Cheng C, Wu E, Kim I, Zhang R, Ma Y, Kortepeter CM, Muñoz MA. Leveraging case narratives to enhance patient age ascertainment from adverse event reports. *Pharm Med*. 2021; Sep; 35(5):307-316. <https://doi.org/10.1007/s40290-021-00398-5> (Journal abstract)
- 26 Dang V, Wu E, Kortepeter CM, Phan M, Zhang R, Ma Y, Muñoz MA. Evaluation of a natural language processing tool for extracting gender, weight, ethnicity, and race in the US food and drug administration adverse event reporting system. *Front Drug Saf Regul*. 2022; Nov 14; 2:1020943. <https://doi.org/10.3389/fdsfr.2022.1020943> (Journal full text)
- 27 Abatamarco D, Perera S, Bao SH, Vermeer P, Botsis T. Training augmented intelligent capabilities for pharmacovigilance: applying deep-learning approaches to individual case safety report processing. *Pharm Med*. 2018; 32(6):391-401. <https://doi.org/10.1007/s40290-018-0251-9> (Journal full text)
- 28 Römning H-J, Pushparajan R. AI translation assistant for pharmacovigilance. Poster presented at: DIA Europe 2021; 2021; Virtual conference. (Webpage accessed 21 March 2025)
- 29 Ghosh, R., Kempf, D., Pufko, A. et al. Automation Opportunities in Pharmacovigilance: An Industry Survey. *Pharm Med* 34, 7-18 (2020). <https://doi.org/10.1007/s40290-019-00320-0> (Journal full text)

- 30 van der Aalst WMP, Bichler M, Heinzl A. Robotic process automation. *Bus Inf Syst Eng.* 2018;60(4):269-272. <https://doi.org/10.1007/s12599-018-0542-4> (Journal full text)
- 31 TransCelerate BioPharma Inc. Successful implementation of robotic process automation (RPA) in the individual case safety report (ICSR) management process leads to enhanced ability to protect patients: featured solution – interactive (ICSR) & automation technologies tool (IATT). [Internet]. Philadelphia (PA): TransCelerate BioPharma Inc.; 2023. (Webpage accessed 23 September 2025)
- 32 Gosselt HR, Bazelmans EA, Lieber T, van Hunsel FP, Härmark L. Development of a multivariate prediction model to identify individual case safety reports which require clinical review. *Pharmacoepidemiol Drug Saf.* 2022;Dec;31(12):1300-1307. <https://doi.org/10.1002/pds.5553> (Journal full text)
- 33 Bergman E, Dürlich L, Arthursen V, Sundström A, Larsson M, Bhuiyan S, Jakobsson A, Westman G. BERT based natural language processing for triage of adverse drug reaction reports shows close to human-level performance. *PLOS Digital Health.* 2023;Dec 6;2(12):e0000409. <https://doi.org/10.1371/journal.pdig.0000409> (Journal full text)
- 34 Cherkas Y, Ide J, van Stekelenborg J. Leveraging machine learning to facilitate individual case causality assessment of adverse drug reactions. *Drug Saf.* 2022;May;45(5):571-582. <https://doi.org/10.1007/s40264-022-01163-6> (Journal abstract)
- 35 Meldau EL, Bista S, Melgarejo-González C, Norén GN. Automated redaction of names in adverse event reports using transformer-based neural networks. *BMC Med Inform Decis Mak.* 2024;Dec23;24(1):401. <https://doi.org/10.1186/s12911-024-02785-9> (Journal full text)
- 36 Finney DJ. Systematic Signalling of Adverse Reactions to Drugs. *Methods Inf Med.* 1974;13(01):1-10. <https://doi.org/10.1055/s-0038-1636131> (Journal full text)
- 37 Bate A, Lindquist M, Edwards IR, Olsson S, Orre R, Lansner A, De Freitas RM. A Bayesian neural network method for adverse drug reaction signal generation. *Eur J Clin Pharmacol.* 1998;Jul;54:315-321. <https://doi.org/10.1007/s002280050466> (Journal abstract)
- 38 DuMouchel W. Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. *Am Stat.* 1999;1;53(3):177-190. <https://doi.org/10.1080/00031305.1999.10474456> (Journal full text)
- 39 Evans SJ, Waller PC, Davis S. Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiol Drug Saf.* 2001;Oct;10(6):483-486. <https://doi.org/10.1002/pds.677> (Journal full text)
- 40 Council for International Organizations of Medical Sciences (CIOMS). Practical aspects of signal detection in pharmacovigilance: CIOMS Working Group VIII report. Geneva: Council for International Organizations of Medical Sciences; 2010. (Full text accessed 15 October 2025)
- 41 Caster O, Norén GN, Madigan D, Bate A. Large-scale regression-based pattern discovery: the example of screening the WHO global drug safety database. *Stat Anal Data Min ASA Data Sci J.* 2010;Aug;3(4):197-208. <https://doi.org/10.1002/sam.10078> (Journal abstract)
- 42 Harpaz R, Perez H, Chase HS, Rabadan R, Hripscak G, Friedman C. Biclustering of adverse drug events in the FDA's spontaneous reporting system. *Clin Pharmacol Ther.* 2011;Feb;89(2):243-250. <https://doi.org/10.1038/clpt.2010.285> (Journal full text)
- 43 Daneshjou R, Altman RB. Data-driven prediction of drug effects and interactions. *Science Translational Medicine.* 2012;Mar14;4(125):125ra31. <https://doi:10.1126/scitranslmed.3003377> (Journal abstract)
- 44 van Puijenbroek E, Egberts A, Heerdink E, Leufkens H. Detecting drug–drug interactions using a database for spontaneous adverse drug reactions: an example with diuretics and non-steroidal anti-inflammatory drugs. *Eur J Clin Pharmacol.* 2000;56(9-10):733-738. <https://doi.org/10.1007/s002280000215> (Journal full text)
- 45 Thakrar BT, Grundschober SB, Doesseger L. Detecting signals of drug–drug interactions in a spontaneous reports database. *Br J Clin Pharmacol.* 2007;Oct;64(4):489-495. <https://doi.org/10.1111/j.1365-2125.2007.02900.x> (Journal full text)
- 46 Norén GN, Sundberg R, Bate A, Edwards IR. A statistical methodology for drug–drug interaction surveillance. *Stat in Med.* 2008;Jul 20;27(16):3057-3070. <https://doi.org/10.1002/sim.3247> (Journal full text)
- 47 Tatonetti NP, Ye PP, Daneshjou R, Altman RB. Data-driven prediction of drug effects and interactions. *Science Translational Medicine.* 2012;Mar14;4(125):125ra31. <https://doi:10.1126/scitranslmed.3003377> (Journal abstract)
- 48 Hauben M. Artificial intelligence and data mining for the pharmacovigilance of drug–drug interactions. *Clin Ther.* 2023;Feb1;45(2):117-133. <https://doi.org/10.1016/j.clinthera.2023.01.002> (Journal full text)
- 49 Hauben M, Hung E, Chen Y. Potential signals of COVID-19 as an effect modifier of adverse drug reactions. *Clin Ther.* 2024;Jan1;46(1):20-29. <https://doi.org/10.1016/j.clinthera.2023.10.002> (Journal full text)

- 50 Sandberg L, Taavola H, Aoki Y, Chandler R, Norén GN. Risk factor considerations in statistical signal detection: using subgroup disproportionality to uncover risk groups for adverse drug reactions in VigiBase. *Drug Saf.* 2020;Oct;43(10):999-1009. <https://doi.org/10.1007/s40264-020-00957-w> (Journal full text)
- 51 Juhlin K, Karimi G, André M, Lucas C, Star K, Norén GN. Using VigiBase to identify substandard medicines: detection capacity and key prerequisites. *Drug Saf.* 2015;38(4):373-382. <https://doi.org/10.1007/s40264-015-0271-2> (Journal full text)
- 52 Trippe ZA, Brendani B, Meier C, de Jong H, Norén GN, Juhlin K. Identification of substandard medicines via disproportionality analysis of individual case safety reports. *Drug Saf.* 2017;40(4):293-303. <https://doi.org/10.1007/s40264-016-0499-5> (Journal abstract)
- 53 Mahaux O, Bauchau V, Zeinoun Z, Van Holle L. Tree-based scan statistic – application in manufacturing-related safety signal detection. *Vaccine.* 2019;37(1):49-55. <https://doi.org/10.1016/j.vaccine.2018.11.044>. (Journal full text).
- 54 Caster O, Juhlin K, Watson S, Norén GN. Improved statistical signal detection in pharmacovigilance by combining multiple strength-of-evidence aspects in vigiRank: retrospective evaluation against emerging safety signals. *Drug Saf.* 2014;Aug;37:617-628. <https://doi.org/10.1007/s40264-014-0204-5> (Journal full text)
- 55 Van Holle L, Bauchau V. Use of logistic regression to combine two causality criteria for signal detection in vaccine spontaneous report data. *Drug Saf.* 2014;Dec;37:1047-1057. <https://doi.org/10.1007/s40264-014-0237-9> (Journal full text)
- 56 Scholl JH, van Hunsel FP, Hak E, van Puijenbroek EP. A prediction model-based algorithm for computer- assisted database screening of adverse drug reactions in the Netherlands. *Pharmacoepidemiol Drug Saf.* 2018;Feb;27(2):199-205. <https://doi.org/10.1002/pds.4364> (Journal full text)
- 57 Ly T, Pamer C, Dang O, Brajovic S, Haider S, Botsis T, Milward D, Winter A, Lu S, Ball R. Evaluation of Natural Language Processing (NLP) systems to annotate drug product labeling with MedDRA terminology. *J Biomed Inform.* 2018;Jul1;83:73-86. <https://doi.org/10.1016/j.jbi.2018.05.019> (Journal full text)
- 58 Shetty KD, Dalal SR. Using information mining of the medical literature to improve drug safety. *J Am Med Inform Assoc.* 2011;Sep1;18(5):668-674. <https://doi.org/10.1136/amiajnl-2011-000096> (Journal abstract)
- 59 Avillach P, Dufour JC, Diallo G, Salvo F, Joubert M, Thiessard F, Mouglin F, Trifirò G, Fourrier-Réglat A, Pariente A, Fieschi M. Design and validation of an automated method to detect known adverse drug reactions in MEDLINE: a contribution from the EU-ADR project. *J Am Med Inform Assoc.* 2013;May1;20(3):446-452. <https://doi.org/10.1136/amiajnl-2012-001083> (Journal abstract)
- 60 Silverman AL, Sushil M, Bhasuran B, Ludwig D, Buchanan J, Racz R, Parakala M, El-Kamary S, Ahima O, Belov A, Choi L. Algorithmic Identification of Treatment-Emergent Adverse Events From Clinical Notes Using Large Language Models: A Pilot Study in Inflammatory Bowel Disease. *Clin Pharmacol Ther.* 2024;Jun;115(6):1391-1399. <https://doi.org/10.1002/cpt.3226> (Journal full text)
- 61 Wu J, Ruan X, McNeer E, Rossow KM, Choi L. Developing a natural language processing system using transformer-based models for adverse drug event detection in electronic health records. *medRxiv.* 2024;Jul10:2024. <https://doi.org/10.1101/2024.07.09.24310100> (Journal full text)
- 62 Tatonetti NP, Denny JC, Murphy SN, Fernald GH, Krishnan G, Castro V, Yue P, Tsau PS, Kohane I, Roden DM, Altman RB. Detecting drug interactions from adverse-event reports: interaction between paroxetine and pravastatin increases blood glucose levels. *Clinical Pharmacology & Therapeutics.* 2011 Jul;90(1):133-142. Erratum in: *Clin Pharmacol Ther.* 2011;Sep;90(3):480. Tsau, P S (corrected to Tsao, P S). <https://doi.org/10.1038/clpt.2011.83> (Journal full text)
- 63 Hauben M, Reynolds R, Caubel P. Deconstructing the pharmacovigilance hype cycle. *Clin Ther.* 2018;Dec1;40(12):1981-1990. <https://doi.org/10.1016/j.clinthera.2018.10.021> (Journal full text)
- 64 Muñoz MA, Dal Pan GJ, Wei YJ, Delcher C, Xiao H, Kortepeter CM, Winterstein AG. Towards automating adverse event review: a prediction model for case report utility. *Drug Saf.* 2020;Apr;43:329-338. <https://doi.org/10.1007/s40264-019-00897-0> (Journal abstract)
- 65 Kreimeyer K, Dang O, Spiker J, Muñoz MA, Rosner G, Ball R, Botsis T. Feature engineering and machine learning for causality assessment in pharmacovigilance: Lessons learned from application to the FDA Adverse Event Reporting System. *Comput Biol Med.* 2021;Aug1;135:104517. <https://doi.org/10.1016/j.combiomed.2021.104517> (Journal abstract)
- 66 Lieber T, Gosselt HR, Kools PC, Kruijssen OC, Van Lierop SNC, Härmark L, van Hunsel F. Natural language processing for automated triage and prioritization of individual case safety reports for case-by-case assessment. *Front Drug Saf Regul.* 2023;3: 1120135. <https://doi.org/10.3389/fdsfr.2023.1120135> (Journal full text)
- 67 European Medicines Agency. 2024 AI Observatory. 6 May 2025. EMA/154528/2025. (Webpage accessed 23 September 2025)
- 68 Singh LL, Sudarsan SD, Jetley RP, Fitzgerald B, Milanova M. Semantic search tool for adverse event reports of medical devices. In: *SWWS 2011:proceedings of the 2011 international conference on semantic web & web services.* 2011;July18-21:129-135. <https://doi.org/10.13140/RG.2.2.22712.29443> (Journal full text)

- 69 Zekarias A, Meldau EL, Bista S, Félix China J, Sandberg L. Narrative Search Engine for Case Series Assessment Supported by Artificial Intelligence Query Suggestions. *Drug Saf.* 2025;Mar15:1-3. <https://doi.org/10.1007/s40264-025-01529-6> (Journal full text)
- 70 Imran M, Bhatti A, King DM, Lerch M, Dietrich J, Doron G, Manlik K. Supervised machine learning-based decision support for signal validation classification. *Drug Saf.* 2022;May;45(5):583-596. <https://doi.org/10.1007/s40264-022-01159-2> (Journal full text)
- 71 Spiker J, Kreimeyer K, Dang O, Boxwell D, Chan V, Cheng C, Gish P, Lardieri A, Wu E, De S, Naidoo J. Information visualization platform for postmarket surveillance decision support. *Drug Saf.* 2020;Sep;43:905-915. <https://doi.org/10.1007/s40264-020-00945-0> (Journal abstract)
- 72 Pinheiro LC, Durand J, Dogné JM. An Application of Machine Learning in Pharmacovigilance: Estimating Likely Patient Genotype from Phenotypical Manifestations of Fluoropyrimidine Toxicity. *Clin Pharmacol Ther.* 2020;107(4):944-947. <https://doi.org/10.1002/cpt.1789> (Journal full text)
- 73 Orre R, Bate A, Norén GN, Swahn E, Arnborg S, Edwards IR. A Bayesian recurrent neural network for unsupervised pattern recognition in large incomplete data sets. *Int J Neural Syst.* 2005;Jun;15(03):207-222. <https://doi.org/10.1142/S0129065705000219> (Journal abstract)
- 74 Ball R, Botsis T. Can network analysis improve pattern recognition among adverse events following immunization reported to VAERS?. *Clin Pharmacol Ther.* 2011;Aug;90(2):271-278. <https://doi.org/10.1038/clpt.2011.119> (Journal abstract)
- 75 Harpaz R, Perez H, Chase HS, Rabadan R, Hripcsak G, Friedman C. Biclustering of adverse drug events in the FDA's spontaneous reporting system. *Clin Pharmacol Ther.* 2011;Feb;89(2):243-250. <https://doi.org/10.1038/clpt.2010.285> (Journal abstract)
- 76 Botsis T, Scott J, Goud R, Toman P, Sutherland A, Ball R. Novel algorithms for improved pattern recognition using the US FDA Adverse Event Network Analyzer. *Ine-Health-For Continuity of Care.* 2014;1178-1182. (Journal full text) <https://doi.org/10.3233/978-1-61499-432-9-1178>
- 77 Fusaroli M, Polizzi S, Menestrina L, Giunchi V, Pellegrini L, Raschi E, Weintraub D, Recanatini M, Castellani G, De Ponti F, Poluzzi E. Unveiling the burden of drug-induced impulsivity: a network analysis of the FDA adverse event reporting system. *Drug Saf.* 2024;Aug15:1-8. <https://doi.org/10.1007/s40264-024-01471-z> (Journal full text)
- 78 Norén GN, Meldau EL, Chandler RE. Consensus clustering for case series identification and adverse event profiles in pharmacovigilance. *Artif Intell Med.* 2021;Dec1;122:102199.
- 79 Okada T, Hashiguchi M, Hori S. Classification of patient characteristics associated with reported adverse drug events to neuraminidase inhibitors: an applicability study of latent class analysis in pharmacovigilance. *Int J Clin Pharm* 2022;44:1332-1341. <https://doi.org/10.1007/s11096-022-01477-6> (Journal full text)
- 80 Bean DM, Wu H, Iqbal E, Dzahini O, Ibrahim ZM, Broadbent M, Stewart R, Dobson RJ. Knowledge graph prediction of unknown adverse drug reactions and validation in electronic health records. *Sci Rep.* 2017;Nov27;7(1):16416. <https://doi.org/10.1038/s41598-017-16674-x> (Journal full text)
- 81 Simms AM, Kanakia A, Sipra M, Dutta B, Southall N. A patient safety knowledge graph supporting vaccine product development. *BMC Med Inform Decis Mak.* 2024;Jan4;24(1):10. <https://doi.org/10.1186/s12911-023-02409-8> (Journal full text)
- 82 Knox C, Wilson M, Klinger CM, Franklin M, Oler E, Wilson A, Pon A, Cox J, Chin NE, Strawbridge SA, Garcia-Patino M. DrugBank 6.0: the DrugBank knowledgebase for 2024. *Nucleic Acids Res.* 2024;Jan5;52(D1):D1265-1275. <https://doi.org/10.1093/nar/gkad976> (Journal full text)
- 83 Kuhn M, Letunic I, Jensen LJ, Bork P. The SIDER database of drugs and side effects. *Nucleic Acids Res.* 2016;Jan4;44(D1):D1075-2079. <https://doi.org/10.1093/nar/gkv1075> (Journal full text)
- 84 Belleau F, Nolin MA, Tourigny N, Rigault P, Morissette J. Bio2RDF: towards a mashup to build bioinformatics knowledge systems. *J Biomed Inform.* 2008;Oct1;41(5):706-716. <https://doi.org/10.1016/j.jbi.2008.03.004> (Journal full text)
- 85 Zhang W, Liu F, Luo L, Zhang J. Predicting drug side effects by multi-label learning and ensemble learning. *BMC Bioinformatics.* 2015;Dec;16:1-1. <https://doi.org/10.1186/s12859-015-0774-y> (Journal full text)
- 86 Low YS, Caster O, Bergvall T, Fourches D, Zang X, Norén GN, Rusyn I, Edwards R, Tropsha A. Cheminformatics-aided pharmacovigilance: application to Stevens-Johnson Syndrome. *J Am Med Inform Assoc.* 2016;Sep1;23(5):968-78. <https://doi.org/10.1093/jamia/ocv127> (Journal full text)
- 87 Bai C, Wu L, Li R, Cao Y, He S, Bo X. Machine learning-enabled drug-induced toxicity prediction. *Adv Sci (Weinh).* 2025;12(12):2413405. <https://doi.org/10.1002/advs.202413405> (Journal full text)
- 88 Hauben M, Rafi M, Abdelaziz I, Hassanzadeh O. Knowledge graphs in pharmacovigilance: a scoping review. *Clin Ther.* 2024;Jul9. <https://doi.org/10.1016/j.clinthera.2024.06.003> (Journal full text)

- 89 Sridharan K, Sivaramakrishnan G. Enhancing readability of USFDA patient communications through large language models: a proof-of-concept study. *Expert Rev Clin Pharmacol*. 2024;Aug2;17(8):731-741. <https://doi.org/10.1080/17512433.2024.23638407> (Journal abstract)
- 90 Ying L, Liu Z, Fang H, Kusko R, Wu L, Harris S, Tong W. Text summarization with ChatGPT for drug labeling documents. *Drug Discov Today*. 2024;May7:104018. <https://doi.org/10.1016/j.drudis.2024.104018> (Journal full text)
- 91 Dowdy K, Hoffman A, Giles T, Kugele D, et al. Summarizing FAERS narratives with generative AI: methods, resource requirements, and quality assessment. In: *FDA Symposium on Scientific Computing and Digital Transformation*; 2024. Silver Spring (MD): Center for Drug Evaluation and Research, U.S. Food and Drug Administration; Booz Allen Hamilton. (Full text accessed 23 September 2025)
- 92 Dietrich J, Hollstein A. Performance and Reproducibility of Large Language Models in Named Entity Recognition: Considerations for the Use in Controlled Environments. *Drug Saf*. 2025;Mar;48(3):287-303. <https://doi.org/10.1007/s40264-024-01499-1> (Journal full text)
- 93 Wu L, Xu J, Thakkar S, Gray M, Qu Y, Li D, Tong W. A framework enabling LLMs into regulatory environment for transparency and trustworthiness and its application to drug labeling document. *Regul Toxicol Pharmacol*. 2024;May1;149:105613 <https://doi.org/10.1016/j.yrtph.2024.105613> (Journal full text)
- 94 Painter JL, Chalamalasetti VR, Kassekert R, Bate A. Automating pharmacovigilance evidence generation: using large language models to produce context-aware structured query language. *JAMIA open*. 2025;Feb;8(1):o0af003. <https://doi.org/10.1093/jamiaopen/oaf003> (Journal full text)
- 95 Benaïche A, Billaut-Laden I, Randriamihaja H, Bertocchio JP. Assessment of the Efficiency of a ChatGPT-Based Tool, MyGenAssist, in an Industry Pharmacovigilance Department for Case Documentation: Cross-Over Study. *J Med Internet Res* 2025;27:e65651. <https://doi:10.2196/65651> (Journal full text)
- 96 Barbieri MA, Battini V, Carnovale C, Cocco M, Papoutsi DG, Heckmann NS, Sessa M. Artificial intelligence in pharmacovigilance signal management: a review of tools, implementations, research, and regulatory landscape. *Expert Opin Drug Saf*. 2025;1-16. <https://doi.org/10.1080/14740338.2025.2545926> (Journal full text)
- 97 European Medicines Agency (EMA). 2024 AI Observatory. 2025 May 6. EMA/154528/2025. (Webpage accessed 23 September 2025)
- 98 Kreimeyer K, Dang O, Spiker J, Gish P, Weintraub J, Wu E, Ball R, Botsis T. Increased confidence in deduplication of drug safety reports with natural language processing of narratives at the us food and drug administration. *Front Drug Saf Regul*. 2022;Jun15;2:918897. <https://doi.org/10.3389/dfsfr.2022.918897> (Journal full text)
- 99 Painter J, Haguinet F, Cranfield C, Bate A. MSR20 NLP and machine learning to automate identification of suspected medication errors from real-world unstructured narratives. *Value Health*. 2023;26(6):S281. <https://doi.org/10.1016/j.jval.2023.03.1556> (Journal full text)
- 100 Caster O, Sandberg L, Bergvall T, Watson S, Norén GN. *vigiRank* for statistical signal detection in pharmacovigilance: first results from prospective real-world use. *Pharmacoepidemiol Drug Saf*. 2017;Aug;26(8):1006-1010. <https://doi.org/10.1002/pds.4247> (Journal full text)
- 101 Scholl JH, van Hunsel FP, Hak E, van Puijenbroek EP. A prediction model-based algorithm for computer-assisted database screening of adverse drug reactions in the Netherlands. *Pharmacoepidemiol Drug Saf*. 2018;Feb;27(2):199-205. <https://doi.org/10.1002/pds.4364> (Journal full text)
- 102 Rudolph A, Mitchell J, Barrett J, Sköld H, Taavola H, Erlanson N, Melgarejo-González C, Yue QY. Global safety monitoring of COVID-19 vaccines: How pharmacovigilance rose to the challenge. *Ther Adv Drug Saf*. 2022;Aug;13:20420986221118972. <https://doi.org/10.1177/20420986221118972> (Journal full text)
- 103 EudraVigilance. European Medicines Agency (EMA). (Webpage accessed 15 October 2025)
- 104 Habets PC, van IJzendoorn DG, Vinkers CH, Härmark L, de Vries LC, Otte WM. Development and validation of a machine-learning algorithm to predict the relevance of scientific articles within the field of teratology. *Reprod Toxicol*. 2022;Oct1;113:150-154. <https://doi.org/10.1016/j.reprotox.2022.09.001> (Journal full text)
- 105 Organisation for Economic Co-operation and Development (OECD). The state of implementation of the OECD AI principles four years on. *OECD Artif Intell Pap*. 2023;(3). Paris: OECD Publishing. <https://doi.org/10.1787/835641c9-en>. (Full text accessed 15 October 2025)
- 106 ArtificialIntelligenceAct.eu. [Internet]. Brussels: European Union; 2024. (Webpage accessed 15 October 2025)
- 107 Government of Canada. Artificial Intelligence and Data Act (AIDA). [Internet]. Ottawa: Government of Canada; 2023. (Webpage accessed 15 October 2025)
- 108 AI Policy Tracker. [Internet]. (Webpage accessed 25 Aug 2025)
- 109 European Medicines Agency (EMA). Reflection paper on the use of artificial intelligence (AI) in the medicinal product lifecycle. Amsterdam: European Medicines Agency; 2024;Sep 9. (Full text accessed 21 March 2025).

- 110 European Commission, Directorate-General for Communications Networks, Content and Technology. The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment. Luxembourg: Publications Office of the European Union; 2020. <https://data.europa.eu/doi/10.2759/002360> (Full text)
- 111 Department of Industry, Science, Energy and Resources (Australia). Australia's artificial intelligence ethics principles. Canberra: Australian Government; 2019. (Webpage accessed 21 March 2025).
- 112 Ministry of Health Singapore. Artificial intelligence in healthcare guidelines (AIHGLE) 1.0. Singapore: Ministry of Health; 2021. (Full text accessed 21 March 2025).
- 113 Government of the United Kingdom, Department for Science, Innovation & Technology. Implementing the UK AI regulatory principles: guidance for regulators. London: Department for Science, Innovation & Technology; 2024. (Full text accessed 21 March 2025).
- 114 United States. Office of Science and Technology Policy. The Blueprint for an AI Bill of Rights: making automated systems work for the American people. [Internet]. Washington (DC): The White House; 2022. (Webpage accessed 15 October 2025)
- 115 Pan American Health Organization (PAHO). Artificial intelligence in public health: digital transformation toolkit, knowledge tools. Washington (DC): Pan American Health Organization; 2021. PAHO/EIH/IS/21-011. (Full text accessed 21 March 2025).
- 116 World Health Organization (WHO). Ethics and governance of artificial intelligence for health: WHO guidance. Geneva: World Health Organization; 2021. Licence: CC BY-NC-SA 3.0 IGO. (Full text accessed 21 March 2025).
- 117 Organisation for Economic Co-operation and Development (OECD). OECD AI principles overview. Paris: Organisation for Economic Co-operation and Development; 2019, updated 2024. (Full text accessed 21 March 2025).
- 118 Heads of Medicines Agencies (HMA), European Medicines Agency (EMA) Joint Big Data Steering Group. Multi-annual artificial intelligence workplan 2023-2028. Amsterdam: European Medicines Agency; 2023. (Full text accessed 21 March 2025).
- 119 U.S. Food and Drug Administration (FDA). Using artificial intelligence and machine learning in the development of drug and biological products. Silver Spring (MD): U.S. Food and Drug Administration; 2023, revised 2025. (Full text accessed 21 March 2025)
- 120 U.S. Food and Drug Administration (FDA). Considerations for the use of artificial intelligence to support regulatory decision-making for drug and biological products: draft guidance for industry and other interested parties. Silver Spring (MD): U.S. Food and Drug Administration; 2025; Jan. (Full text accessed 15 October 2025)
- 121 U.S. Food and Drug Administration (FDA), Center for Drug Evaluation and Research (CDER). Artificial intelligence in drug manufacturing. Silver Spring (MD): U.S. Food and Drug Administration; 2023. (Full text accessed 21 March 2025)
- 122 U.S. Food and Drug Administration (FDA). CDER Emerging Drug Safety Technology Program (EDSTP). Silver Spring (MD): U.S. Food and Drug Administration; 2024. (Full text accessed 21 March 2025)
- 123 European Medicines Agency (EMA). Harnessing AI in medicines regulation: use of large language models (LLMs). 2024. (Full text accessed 21 March 2025)
- 124 Government of Canada. Guide to the use of generative AI. Ottawa: Treasury Board of Canada Secretariat; 2023. (Full text accessed 21 March 2025)
- 125 Organisation for Economic Co-operation and Development (OECD). Initial policy considerations for generative artificial intelligence. Paris: Organisation for Economic Co-operation and Development; 2023. (Full text accessed 21 March 2025)
- 126 World Health Organization (WHO). Ethics and governance of artificial intelligence for health: guidance on large multi-modal models. Geneva: World Health Organization; 2024. Licence: CC BY-NC-SA 3.0 IGO. (Full text accessed 21 March 2025)
- 127 World Health Organization (WHO). Regulatory considerations on artificial intelligence for health. Geneva: World Health Organization; 2023. Licence: CC BY-NC-SA 3.0 IGO. (Full text accessed 21 March 2025)
- 128 European Commission. Ethics guidelines for trustworthy AI. [Internet]. Brussels: European Commission; 2019 (Full text accessed 21 March 2025).
- 129 Organisation for Economic Co-operation and Development (OECD). Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449 (2019, amended 2024). Paris: Organisation for Economic Co-operation and Development; 2024. (Full text accessed 21 March 2025)
- 130 U.S. Food and Drug Administration (FDA), Health Canada, Medicines and Healthcare products Regulatory Agency (MHRA). Good machine learning practice for medical device development: guiding principles. Silver Spring (MD): U.S. Food and Drug Administration; 2021. (Full text accessed 21 March 2025)
- 131 U.S. Food and Drug Administration (FDA). Transparency of machine-learning enabled medical devices: guiding principles. [Internet]. Silver Spring (MD): U.S. Food and Drug Administration; 2022. (Full text accessed 15 October 2025)
- 132 ISO/IEC 23894:2023 Information technology - Artificial intelligence - Guidance on risk management. (Webpage accessed 15 October 2025)

## 第3章

# リスクに基づくアプローチ

### 原則

リスクに基づくアプローチは、AIシステムが引き起こし得る潜在的な危険性を認識し、コアとなるPV業務の実行において、各ユースケースが異なる種類および水準のリスクを伴うことを前提とします。このため、PVシステムの挙動および結果に悪影響を及ぼし得るリスクを、プロセス管理を考慮しながら特定し、優先順位付けし、管理するためのリスク評価の実施が求められます。リスクは、予測される影響の大きさと、負の結果が生じる可能性の双方によって特徴付けられます。<sup>1</sup>

このアプローチはまた、リスクの程度に応じて、誤りおよびバイアスを特定し低減するための手順を支援します。AIソリューションの導入戦略は、原則として特定されたリスクの水準に見合ったものであるべきです。

### 主要メッセージ

- AIをPVプロセスへ統合する際には、AIアルゴリズムおよび人間の双方のパフォーマンスが完全ではないことを考慮する必要があります。
- PVにおけるAI利用に関連して生じ得るリスクは、患者安全、PV利用者の信頼および関与、PVプロセスの効率性、ならびに規制要件および倫理原則への適合性に影響を及ぼす可能性があります。
- 重要度の高い領域に努力と資源を集中させることにより、適切なリスクに基づくアプローチは、患者安全およびPV関係者に不利益を与えることなく、AIの能力を最大限に活用することを可能にします。
- リスクに基づくアプローチは、人間による監督の形態、妥当性と頑健性の戦略、透明性の水準、公正性と衡平性（fairness and equity）の確保、ならびにデータプライバシー維持の取組みに適用されます。
- リスク評価では、AIシステムそのもの、使用状況（context of use）、ならびにリスクが現実化した場合の影響および発生可能性を考慮する必要があります。
- リスクに基づくアプローチは、定期的に、またシステム性能の変化に応じて、必要に応じて見直しおよび適応が行われるべきです。

## 3.1. はじめに

### 3.1.1. 規制上の考慮事項

AI要素の統合の有無にかかわらず、PVシステムは既存の規制および製造販売後安全管理の基準（GVP）に適合することが求められます。<sup>2,3</sup> GVPに従い、個別症例安全性報告（ICSR）の収集および取扱い、シグナルマネジメント、定期的安全性報告を含む広範なPVプロセスは、PVの目標および目的を達成するために重要な活動と位置付けられています。<sup>6</sup> 規制枠組みでは一般に、AIシステムの開発、展開監視、文書化、および規制上の監督においてリスクに基づくアプローチを採用することが推奨されます。

これは、システムライフサイクル全体を通じて関連リスクを予測し、特定し、軽減することを確実にするためです。<sup>4,5,6</sup> EU AI Act<sup>7</sup>では、AIシステムを四つのリスク区分、すなわち低または最小リスク、限定的リスク（透明性義務）、高リスク、および許容不可能リスク（禁止されるAI利用）に分類しています。高リスクAIシステムには、例えばAIベースの医療ソフトウェア／医療機器や、採用活動に用いられるAIシステムなどが含まれ、提供者および利用者に対して厳格な要件および義務が課されます。これには、リスク軽減システムの導入、学習・検証・試験のための高品質データセットの確保、活動ログの記録、詳細な文書化、明確なユーザー情報の提供、人間による監督、ならびに高い頑健性、正確性およびサイバーセキュリティの確保が含まれます。

本報告書で提示される指針原則は、EU AI Actにおける高リスクAIシステムの要件と多くの点で重複していますが、組織のPVプロセスへの統合を検討するAIシステムがどのリスク区分に該当するか判断には、個別事例ごとの慎重な評価が必要となり、必要に応じて法的助言を求めることが想定されます。医薬品ライフサイクルにおいては、EMAは患者安全に影響を与えるユースケースを「高患者リスク（high patient risk）」、規制上の意思決定に重大な影響を及ぼすユースケースを「高規制影響（high regulatory impact）」として位置付けています。<sup>8</sup> また、カナダにおけるAIDAは、EU AI Act、OECD AI Principles、および米国国立標準技術研究所（NIST）のRisk Management Framework（RMF）を含む国際的規範と整合したリスクベースドアプローチに基づき、責任あるAIの開発を確保する目的で策定されました。<sup>9</sup>

AIソリューションの開発段階およびライフサイクルのその他の段階において、申請者および開発者は、個々の患者、公衆衛生、または規制上の意思決定に対するリスク水準に応じて、規制当局と積極的に関与し、適切な科学的助言を求めることを検討すべきです。必要に応じて、医薬品、医療機器および／またはソフトウェア開発に適用される法令または規制要件に基づき、適切な手続きを通じてAI技術の技術的適格性評価（technical qualification）を受けることが求められます。<sup>12,10</sup> AI技術は急速に進展しているため、PVにおけるAIの利用は、この進化する分野への適応と継続的な理解が求められる規制当局にとっても課題となり、<sup>10</sup> 同時に規制遵守を維持する必要がある産業界のPV関係者にとっても課題となります（医薬品安全性監視における人工知能の開発と展開に関する今後の検討については第10章参照）。

### 3.1.2. 他の指針原則との関係および位置づけ

AIシステムをPVプロセスへ統合することは、人為的誤り、不整合、および人間の限界への対応に寄与し得ますが、一方で一定のリスクおよび課題を伴います。適切なリスクに基づくアプローチにより、組織はAIの能力を最大限に活用しつつ、前述した指針原則が確実に維持されるよう、重要度の高い領域に努力および資源を集中させることが可能になります。

リスクに基づくアプローチは、本報告書で提示される他の指針原則にも適用され、またそれらから影響を受けます。特に、リスクに基づくアプローチは、AIを含むPVプロセスにおいて、人間による監督をどこで、いつ、どのように、そしてどの程度実施すべきかを決定する指針となり、その他のリスク低減活動にも影響を与えます。逆に、AIソリューションのリスク評価は、既存の人間による監督の程度および性質を考慮して実施される場合があります（第4章「ヒューマン・オーバーサイト」参照）。さらに、リスクに基づくアプローチは、AIシステムの試験およびバリデーション（第5章「妥当性と頑健性」参照）ならびに文書化および記録保持の水準（第6章「透明性」参照）にも適用されるべきです。また、データプライバシーおよび公正性と衡平性にも関連します。例えば、AIシステムは、特定の集団に不利益を与えたり、十分なサービス提供が行われなくなったり、またはバイアスが生じたりする可能性のあるリスクについて評価されるべきでありそれらのリスクは適切に低減されなければなりません（第8章「公正性と衡平性」参照）。

### 3.1.3. リスクの種類

本節では、PVにおけるAIシステムの利用に関連して生じ得るいくつかのリスクについて、簡潔に概説します。

#### 患者安全および公衆衛生に対するリスク

PVにおいてAIシステムが不適切に使用された場合、またはその性能が不十分である場合、医薬品の有害作用の検出、収集、評価、理解および予防というPVの目的達成が妨げられる可能性があり、その結果として患者安全、公衆衛生、および規制要件への適合に悪影響を及ぼすおそれがあります。偽陰性または偽陽性、不公正なバイアスを含む、AIシステムによって生成される信頼性の低いまたは不正確なアウトプットは、関連する有害事象（AE）が捕捉されない、症例処理において誤分類が生じる、またはシグナルが見逃されるなど、PV活動に悪影響を及ぼす可能性があります。その結果、安全性上の問題が特定されない、あるいは特定が遅延する可能性があり、患者を危険にさらすおそれがあります。まれではありますが、新規かつ予期されない安全性シグナルの検出が遅れることにより、公衆衛生へ重大な影響を及ぼす場合（いわゆる「ブラックスワン事象」）も考えられます。<sup>11</sup> さらに、当初は十分に機能していたAIシステムであっても、モデルドリフトなどにより時間の経過とともに性能が低下したり、ITインシデントやシステム障害により機能停止に至ったりする可能性があり、その場合、当該AIシステムが支援することを目的としていたPV活動が阻害されるおそれがあります。

#### 利用者の信頼および関与に関するリスク

特定のAIアルゴリズムにおける透明性や解釈可能性の欠如、または人間にとって高度な認知判断を要すると認識される業務（例：因果関係評価）へのAI適用は、PV専門家<sup>12</sup>を含む利用者の信頼および受容を妨げる可能性があります（第6章「透明性」参照）。また、妥当性と頑健性が不十分なAIシステムに関する過去の不良経験も、AI全般に対する不信感につながる可能性があります。

臨床現場におけるPVでは、「個別性軽視（uniqueness neglect）」と呼ばれる、患者がより正確なコンピュータよりも人間の臨床医を好む傾向も、信頼低下の要因となり得ます。これは機械が個々の患者の人間の特性を十分に理解できないという認識に起因します。<sup>13</sup> さらに、特定の集団に対する性能低下や、AIソリューションの開発または運用過程において個人情報機密性が適切に保護されない場合も、不信の原因となり得ます。一方で、検証により高い性能が示されたAIソリューションに対して利用者が過度な信頼を置くことにより、自動化バイアスが生じ、誤ったアウトプットを無意識に受け入れてしまうリスクも存在します。加えて、既存の業務フローやシステムへのAI導入は、技術的、組織的、文化的課題を伴う可能性があり、適切な教育訓練および変更管理戦略が不足している場合には、職務意欲や満足度の低下を招くリスクもあります（第4章「ヒューマン・オーバーサイト」参照）。

#### 効率性に対するリスク

PVプロセスへのAI統合は一般に効率向上を目的としていますが、質の低いAIソリューションは、かえって作業負担を増大させる可能性があります。例えば、AIの出力内容を理解し検証するため、あるいは許容可能な水準に修正するために多大な時間を要する場合、削減できるはずの作業量以上の手作業が発生する可能性があります。また、AI出力に基づく解釈や対応

に不確実性（例：偽陽性）が含まれる場合、限られた資源の非効率的または最適とは言えない利用につながる可能性があります。さらに、すべてのPV上の課題にAIソリューションが必要であるとは限らない点を認識することも重要です。

## その他のリスクの種類

その他のリスクとして、不適切な目的との不整合または誤用、<sup>14</sup>ならびにデータプライバシー、サイバーセキュリティ、知的財産、責任所在、経済的影響および評判に関するリスクが挙げられます。

本章の残りの部分では、主としてAIシステムのPVプロセスへの利用が患者安全に及ぼし得る影響に焦点を当てます。その他のリスクおよび課題については、第7章「データプライバシー」第8章「公正性と衡平性」第6章「透明性」第4章「人間による監視」第9章「ガバナンスとアカウントビリティ」の各章でさらに詳しく論じます。

## 3.2. リスクアセスメント

### 3.2.1. 一般的考慮事項

PVプロセスを支援する目的でAIの導入を計画する組織は、包括的なリスク分析を実施することが求められます。この評価は各AIシステムごとに実施されるべきであり、開発段階から日常運用に至るまで、AIソリューションのライフサイクル全体に適用されるリスク比例型アプローチの基盤となるものです。

PVシステムにおけるAI導入に関連するリスクレベルを判断する際には、AI技術そのもの、使用文脈（context of use）、リスクが実際に発現する可能性、その検出可能性、および潜在的影響が主要な検討要素となります。

### 人工知能技術

リスクの程度は、使用されるシステムの種類（例：静的モデルか動的モデルか）、基盤となるデータ（種類および品質）、技術の新規性（すなわち、従来技術の方がリスク特性がより明確に把握されている可能性）、またはシステムの成熟度（すなわちライフサイクル段階）に依存する場合があります。

特に、PVプロセスへの生成AI（GenAI）モデルの統合に際しては慎重な対応が求められます。より単純または説明可能性の高いAI手法と比較して、GenAIおよび類似モデルは非決定論的な性質を有し、学習データの不透明性やハルシネーション（人間にとって一見整合的かつ説得力のある出力を生成するものの、実際には誤解を招く可能性のある出力）の発生可能性があるため、問題の検出および軽減がより困難となる場合があります。このため、追加的なガードレールの検討が必要となる可能性があります（リスク軽減の節を参照）。また、マルチエージェントシステムは、個々のエージェントに付与される自律性や、エージェント間依存関係に起因する連鎖的障害（cascading failures）に関連した固有のリスクを伴う可能性があります。<sup>15</sup>

AIを取り巻く環境が継続的に進化するにつれ、AI関連リスク領域も同様に変化していきます。新たなリスクが出現する一方で、GenAI/LLMに関連する現在の課題を含め、既存の問題が将来的に解決される可能性もあります。

## 利用状況および影響度

これらは、PVシステム全体の中におけるAIソリューションの位置付けおよび重要性を広く指し、具体的には以下を含みます。

- AIソリューションが、重要なPVプロセスまたは高リスク文脈において使用されるかどうか（例：公衆衛生上の緊急使用、新規有効成分、治験症例など）
- 特定のプロセス内のどの段階でAIシステムが介入するか（例：シグナルレビュー前段階としての関連症例の自動トリアージ）、および当該ソリューションが補助的（assistive）な役割を担うのか、あるいはPVプロセスを直接支援するものか
- 意思決定において、モデル出力が他の情報源または活動と比較してどの程度重要であるか
- 当該プロセスにおける人間の関与と監督の程度（第4章「ヒューマン・オーバーサイト」参照）

## 影響と発生可能性

システムの不具合やモデル性能の不十分さが発生した場合でも、その発生確率および影響の大きさは一様ではありません。例えば、非常に大規模なデータベースに適用される重複検出ソリューションでは、重複症例の100%検出は必ずしも想定されておらず、検出漏れが生じたとしても患者安全の観点では影響がない、または限定的である場合があります。一方で、大規模な患者曝露状況において極めて重大なシグナルの検出が遅延する事象は発生頻度こそ非常に低いものの、公衆衛生に重大な影響を及ぼす可能性があります（いわゆるブラックスワン事象）。

### 3.2.2. 構造化アプローチの例

リスクに基づく評価フレームワークは、さまざまな分野において提案されており、PVシステムにAIソリューションを導入しようとする組織にとって参考となり得ます。ここでは、代表的な例を簡潔に紹介します。

## 信頼性評価フレームワーク

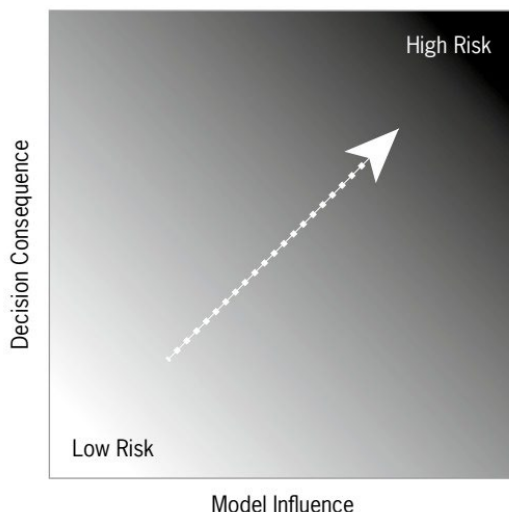
米国FDAは、医薬品の安全性、有効性、または品質に関する規制上の意思決定を支援することを目的として、AIモデルが生成する情報またはデータの信頼性を実証するための段階的アプローチを提案しています（第2章「ランドスケープ分析」参照）。<sup>16</sup> 同様のフレームワークは、医療機器申請における計算モデルの利用<sup>16</sup>や医薬品開発においても提案されています。<sup>17</sup> 以下に示す信頼性評価の初期ステップは、モデルリスクの評価に役立ちます。

1. **検討対象となる問いの定義（Define the question of interest）** AIモデルによって対応すべき具体的な問い、意思決定、または懸念事項を記述します。
2. **使用目的の定義（Define the context of use）** 関心対象の問いに対してモデルがどのように使用されるかを記述します。すなわち、何をモデル化するのか、モデル出力がどのように利用されるのか、またモデル出力と併せて他の情報が使用されるかを明確にします。
3. **AIモデルリスクの評価（Assess the AI model risk）** モデルリスクは以下の二要素により定義されます。(i) 関心対象の問いを判断する際に、AIモデルから得られる証拠が他の証拠と比較してどの程度寄与するか（モデル影響度：model influence）(ii) 関心対象

に関する誤った意思決定によって生じる望ましくない結果の重大性（意思決定結果：decision consequence）意思決定結果およびモデル影響度の評価はそれぞれ独立して決定されますが、使用目的の内容によって規定されるため、モデルリスクはケースごとに異なるものとなります。AIモデルリスク評価には、専門的知識を有する担当者の関与が必要です。Figure 4 に示されるように、意思決定結果またはモデル影響度が増大するにつれて、モデルリスクは低リスクから高リスクへと移行します。

#### 図4：モデルのリスクマトリックス

出典：U.S. Food and Drug Administration. 161



#### アルゴリズム影響評価

カナダでは、アルゴリズム影響評価（Algorithmic Impact Assessment：AIA）ツール<sup>18</sup>が、自動化された意思決定システムに関連するリスクを各省庁および機関がより適切に理解し管理できるよう設計されています。このツールは、多数の質問項目から構成されており、リスク領域（すなわち、プロジェクト、システム設計、アルゴリズム、意思決定、影響、およびデータ）ならびに緩和領域（すなわち、協議、リスク低減および緩和措置）に関する要素を評価します。これらの回答結果に基づいて評価スコアが算出されます。各質問の値は、自動化プロジェクトにおいて導入または軽減されるリスクの程度に応じて重み付けされています。リスク領域では65項目の質問があり、最大スコアは169点です。緩和領域では41項目の質問があり、最大スコアは75点となっています。算出されたスコアの割合に基づき、自動化意思決定システムの影響レベルは以下の4段階に分類されます。

- レベル I – 影響はほとんどない、または影響なし（0%～25%）
- レベル II – 中程度の影響（26%～50%）
- レベル III – 高い影響（51%～75%）
- レベル IV – レベル IV：非常に高い影響（76%～100%）

自動化意思決定指令（Directive on Automated Decision-Making）に基づき、いかなる自動化意思決定システムも本番運用に移行する前に、このアルゴリズム影響評価の実施が求められます。<sup>19</sup>

### 3.3. 問題の検出とリスクの軽減

#### 継続的モニタリングによる課題検出

リスクをいつ低減すべきかを定義するためには、AIシステムの開発段階で策定される、あらかじめ定義されたリスク比例型の試験および検証計画に基づき、どのように課題を検出するかを把握しておく必要があります。試験および検証は、AIシステムの成熟度に応じた異なるレベルを考慮するコンピュータ化システムバリデーション（CSV）の重要なステップです。製薬企業および規制当局に広く採用されている国際製薬技術協会（ISPE）のGAMP 5最新版には、AIおよび機械学習（ML）に焦点を当てた付録が含まれています。<sup>20</sup> 試験は、事前に定義された主要業績評価指標（KPI）および受入基準に基づいて実施されるべきであり、人間のパフォーマンスも考慮するとともに、例えば低品質データなど、特定されたリスク領域を反映する必要があります。

AIシステムが目的適合性（fit for purpose）を満たすことが確認され導入された後は、その性能を継続的に監視し、課題が検出された場合にはリスク低減措置を発動できる継続的プロセスを確立する必要があります。

リスクに基づくアプローチは、導入初期段階では非常に保守的に運用される場合があります、例えばヒューマン・イン・ザ・ループ（HITL）の高い割合など、あらかじめ定められた追加的な低減措置が講じられることがあります。事前に定義された指標および専門家によるサンプル出力の評価に基づき、日常運用における性能への信頼が時間とともに高まった場合には、人間による統制の頻度、量（例：サンプル数）、または深度を段階的に低減することが検討されます。さらに主要AIソリューションの監視を支援するためにAIモデルを用いる、すなわちAI支援型ヒューマン・オーバーサイトも検討可能です（第4章「ヒューマン・オーバーサイト」参照）。

#### 事後的リスク低減アプローチ

課題または性能逸脱が検出された場合、リスクベースに基づく低減措置には、以下が含まれる場合があります。

- **ヒューマン・イン・ザ・ループ（Human-in-the-Loop：HITL）**：人間によるレビューまたは品質管理（QC）を増加させる、あるいは全面的に実施します。これは無期限、または性能が受入基準内に回復するまで継続されます。例えば、重篤性判定アルゴリズムが一部症例において重篤性基準を検出できない（すなわち偽陰性が発生する）場合には、問題が理解され是正されるまで、非重篤と分類されたすべての症例を人間がレビューすることが低減措置となり得ます。
- **モデル再学習（Model re-training）**：最近のデータまたは困難な事例を用いて、基盤モデルに対する対象を絞った再学習を実施します。
- **システムの廃止（Decommissioning）**：低減措置が非効率または過度に高コストであると判断される場合には、システムを停止します。その場合には代替アプローチを検討する必要があります。

#### その他（予防的）リスク低減アプローチ

- **LLM特有の対策**：検索拡張生成（Retrieval Augmented Generation：RAG）などのグラウンディング技術、ハルシネーションを防止するための各種ガードレール<sup>21</sup>、またはマルチエージェントシステムの障害に対応するためのコンティンジェンシープロトコルを含みます。

- **AI出力の不確実性または信頼度の明示**：AI出力に対して、不確実性の水準または信頼度スコアを明確に示すアプローチです。<sup>16</sup>
- **オートメーションバイアスまたは過信への対策**：<sup>22</sup>例えば、模擬データシミュレーションや、訓練環境において意図的に偽陽性出力を注入し、検証・評価を行う方法などが含まれます。
- **レッドチーミング**：特に高リスク状況において適用される手法です。レッドチーミングとは、組織のセキュリティ体制に対して潜在的な攻撃者の能力を模倣することを目的として、権限を付与された専門チームが攻撃または悪用を再現する活動を指します。その目的は、攻撃が成功した場合の影響を明確化するとともに、防御側（Blue Team）が実運用環境において有効に機能する対策を検証し、組織のサイバーセキュリティを強化することにあります。これは「Cyber Red Team」とも呼ばれます。<sup>23</sup>

上記の事項については、第5章「妥当性と頑健性」、第4章「人間による監視」、第9章「ガバナンスとアカウントビリティ」の各章においてさらに詳述されています。

### 3.4. リスクに基づくアプローチのレビューと文書化

リスクに基づくアプローチは、あらかじめ定められた定期的な間隔で見直されるとともに、AIソリューションに性能上の問題が認められた場合には、必要に応じて監督措置を適切に調整する必要があります。AI技術の進化および新たなリスク低減技術の登場を踏まえると、リスク評価フレームワークもまた、動的かつ適応可能な形で運用されることが求められます。

さらに、特に重要なPVプロセスに導入されているAIコンポーネントについては、組織の事業継続計画（Business Continuity Plan：BCP）に組み込む必要があります。その目的は、AIソリューションの障害または性能低下が発生した場合でも、PVシステムの目的達成および規制遵守を維持することにあります。

AIに関連するリスクマネジメント戦略の主要構成要素は、文書化される必要があります（第6章「透明性」参照）。これには、以下が含まれます。

- AIシステムのリスク評価
- KPI、受入基準、および試験・バリデーション活動（比較評価を含む）の結果を含む試験計画
- ヒューマン・オーバーサイト戦略、QC（品質管理）の強化または緩和の基準、ならびに導入後の継続的モニタリングを含むリスク低減措置計画
- リスクマネジメント戦略の定期的再評価および更新計画
- 事業継続計画（Business Continuity Plan）

## 第3章 – 参考文献

- 1 Council for International Organizations of Medical Sciences (CIOMS). Practical approaches to risk minimisation for medicinal products. Geneva: Council for International Organizations of Medical Sciences; 2014. ([Full text](#) accessed 21 March 2025)
- 2 European Medicines Agency (EMA). Guideline on good pharmacovigilance practices (GVP): Module I – pharmacovigilance systems and their quality systems. London: European Medicines Agency; 2012. ([Full text](#) accessed 21 March 2025)

- 3 U.S. Food and Drug Administration (FDA). Good pharmacovigilance practices and pharmacoepidemiologic assessment: guidance for industry. Silver Spring (MD): U.S. Food and Drug Administration; 2005;Mar. ([Full text](#) accessed 21 March 2025).
- 4 World Health Organization (WHO). Regulatory considerations on artificial intelligence for health. Geneva: World Health Organization; 2023. Licence: CC BY-NC-SA 3.0 IGO. ([Full text](#) accessed 21 March 2025)
- 5 European Medicines Agency (EMA). Reflection paper on the use of artificial intelligence (AI) in the medicinal product lifecycle. Amsterdam: European Medicines Agency; 2024;Sep9. ([Webpage](#) accessed 21 March 2025)
- 6 U.S. Food and Drug Administration (FDA). Considerations for the use of artificial intelligence to support regulatory decision-making for drug and biological products: draft guidance for industry and other interested parties. Silver Spring (MD): U.S. Food and Drug Administration; 2025;Jan. ([Full text](#) accessed 21 March 2025)
- 7 European Parliament, Council of the European Union. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828. Off J Eur Union. 2024;L 1689:1-159. ([Webpage](#) accessed 21 March 2025)
- 8 European Medicines Agency (EMA). Reflection paper on the use of artificial intelligence (AI) in the medicinal product lifecycle. Amsterdam: European Medicines Agency; 2024;Sep9. ([Webpage](#) accessed 21 March 2025)
- 9 Government of Canada. The Artificial Intelligence and Data Act (AIDA) – companion document. [Internet]. Ottawa: Government of Canada; 2022. ([Webpage](#) accessed 21 March 2025)
- 10 Hines PA, Herold R, Pinheiro L, Frias Z, Arlett P. Artificial intelligence in European medicines regulation. Nat Rev Drug Discov. 2023;Feb;22(2):81-82. <https://doi:10.1038/d41573-022-00190-3>. ([Journal full text](#))
- 11 Kjoersvik O, Bate A. Black Swan Events and Intelligent Automation for Routine Safety Surveillance. Drug Saf. 2022;May;45(5):419-427. <https://doi:10.1007/s40264-022-01169-0>. ([Journal full text](#))
- 12 Ball R, Talal AH, Dang O, Muñoz M, Markatou M. Trust but verify: lessons learned for the application of AI to case-based clinical decision-making from postmarketing drug safety assessment at the US Food and Drug Administration. J Med Internet Res. 2024;Jun 6;26:e50274. <https://doi:10.2196/50274> ([Journal full text](#))
- 13 Hauben M. Artificial intelligence in pharmacovigilance: Do we need explainability? Pharmacoepidemiol. Drug Saf. 2022;Dec31(12):1311-1316. [https://doi: 10.1002/pds.5501](https://doi:10.1002/pds.5501). ([Journal full text](#))
- 14 Department for Science, Innovation and Technology (DSIT), AI Safety Institute. International AI Safety Report 2025. London: UK Government; 2025;Jan29. ([Webpage](#) accessed 18 September 2025)
- 15 Siva Kumar RS. Taxonomy of failure modes in agentic AI systems. [Internet]. Redmond (WA): Microsoft Security Blog; 2025;Apr24. ([Webpage](#) accessed 18 September 2025)
- 16 U.S. Food and Drug Administration (FDA). Assessing the credibility of computational modeling and simulation in medical device submissions. Silver Spring (MD): U.S. Food and Drug Administration; 2023. ([Full text](#) accessed 21 March 2025)
- 17 International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). General principles for model-informed drug development (M15): draft version endorsed on 6 November 2024. Geneva: ICH; 2024. ([Full text](#) accessed 21 March 2025)
- 18 Government of Canada. Algorithmic Impact Assessment tool. [Internet]. Ottawa: Government of Canada; 2020 ([Webpage](#) accessed 21 March 2025)
- 19 Government of Canada. Directive on automated decision-making. [Internet]. Ottawa: Government of Canada; 2019 ([Webpage](#) accessed 21 March 2025)
- 20 International Society for Pharmaceutical Engineering (ISPE). GAMP 5 guide: a risk-based approach to compliant GxP computerized systems. 2nd ed. Tampa (FL): International Society for Pharmaceutical Engineering; 2022. ([Abstract](#) accessed 21 March 2025).
- 21 Hakim JB, Painter JL, Ramcharran D, Kara V, Powell G, Sobczak P, Sato C, Bate A, Beam A. The need for guardrails with large language models in pharmacovigilance and other medical safety critical settings. Sci Rep. 2025;Jul 31;15(1):27886. <https://doi.org/10.1038/s41598-025-09138-0> ([Journal full text](#))
- 22 Adler-Milstein J, Redelmeier DA, Wachter RM. The limits of clinician vigilance as an AI safety bulwark. JAMA. 2024;331(14):1173-1174. <https://doi:10.1001/jama.2024.3620> ([Journal full text](#))
- 23 National Institute of Standards and Technology (NIST). Glossary. [Internet]. Gaithersburg (MD): NIST Computer Security Resource Center; 2025. ([Webpage](#) accessed 23 October 2025)

## 第4章

# ヒューマン・オーバーサイト

### 原則

ヒューマン・オーバーサイト（人間による監視・監督、人的監視）とは、医薬品安全性監視（PV）におけるAIシステムの設計、実装、監視、および分析において、人間に期待される役割を指します。これは、AIソリューションの性能を管理し、それに関連する潜在的な問題を検出し、軽減するための枠組みを必要とします。

### 主要メッセージ

- ヒューマン・オーバーサイトは、PVに導入されたAIシステムの性能最適化を支援し、信頼性および説明責任を高めます。
- AIシステムに対するヒューマン・オーバーサイトの範囲および性質は、リスクに基づくアプローチに従うべきです。
- 品質保証（QA）の原則は、PVにおけるAIソリューションに対するヒューマン・オーバーサイトの実施にも適用されるべきです。
- PVプロセスを支援する自動化およびAIの利用拡大に伴い、AIと人間の専門性を統合するためのスキルセットの再定義が求められます。これは、意思決定プロセスにおける頑健性および信頼性を確保するためです。その結果、従来の役割および能力は変化し、適切なチェンジマネジメントおよび教育・訓練戦略が必要となります。

## 4.1. はじめに

### 4.1.1. 背景および目的

ヒューマン・オーバーサイトは、AIソリューションが人間の自律性を損なったり、その他の望ましくない、または意図しない影響を引き起こしたりするリスクを最小化するために必要です。<sup>1</sup> 人間の自律性の保護という原則は、人間がAIシステムに対する統制を維持することを求めています。<sup>2</sup> 人間の主体性およびヒューマン・オーバーサイトは、Assessment List for Trustworthy Artificial Intelligence（ALTAI）、EU AI Act、カナダのAIDAなど、複数の規制枠組みにおいて、特に高リスクシステムに対する信頼できるAIの重要な要件とされています<sup>4,3,4</sup>（第2章「ランドスケープ分析」参照）。人間によるレビューのみでは出力の完全な正確性が保証されるわけではありませんが、AIシステムの性能を監視し、必要に応じて修正を行うためにはヒューマン・オーバーサイトが不可欠であり、AIシステムの応用、特に一部の高リスクAIシステムの応用においてはとりわけ、信頼性および人間の説明責任が高まります。

AIシステムは、多くの場合、人間が行ってきた手作業で労力を要する、あるいは複雑な作業を削減すること、またはインテリジェンス拡張ツールとして人間の能力を強化することを目的としています。しかしながら、特定のPV業務の複雑性および重要性、さらにPVデータの複雑かつ変動性の高い性質により、AIコンポーネントは性能が向上していく一方で、依然として不完全な結果を示す可能性があります。そのため、一部のPV向けAI

システムでは、開発、評価、および導入の各段階において、リスクを監視および低減するため、より広範な人間の関与が必要となる場合があります。AIに対する品質保証（QA）アプローチを定義するうえでの重要な出発点かつ主要な課題は、AIソリューションがもたらす効率向上と、高品質な出力を確保するために必要となる人間の介入水準との適切なバランスを見出すことです。端的に言えば、理想的には人間の専門家は機械がうまくできる仕事をすべきではなく、機械は人間の専門家がうまくできる仕事を適切に行えないまま稼働すべきではありません。<sup>5</sup>

ヒューマン・オーバーサイトは、健全なリスクベースアプローチの根幹を成します（第3章「リスクに基づくアプローチ」参照）。AIシステムによる誤りや不適切な出力が見逃された場合に生じ得る影響の大きさに応じて、人間によるAIソリューションの性能監視の水準は比例的に設定されるべきです。

## 4.2. 人間の関与と監督に関する考慮事項

### 4.2.1. 学際的専門性

AIシステムをPVシステムへ適切に統合するためには、開発段階から日常運用に至るまでソリューションのライフサイクル全体を通じて、必要に応じた学際的な人材の専門知識を動員することが求められます。この学際的専門性は通常、分野横断的な専門家間の緊密な協働によって確保されます。該当する場合には、PV専門家、QAスタッフ、データサイエンティスト、統計家、AI/MLエンジニア、データエンジニア、プロンプトエンジニア、IT専門家、サイバーセキュリティの専門家、プラットフォームアナリスト、ソフトウェアエンジニア、倫理専門家、法律専門家、データ保護責任者、プロジェクトマネージャ、上位管理職などが含まれます（第5章「妥当性と頑健性」、第9章「ガバナンスとアカウントビリティ」参照）。

PV専門職、すなわちICSR管理、シグナル検出および分析、またはリスク管理における中核業務を担う担当者は、強固な「ドメイン専門性」または「主題専門性」を有しておりこれはAI機能をPVプロセスへ効果的に統合するうえで不可欠です。そのため、PV専門職は、AIシステムが目的適合性を満たし、かつ最終的な利用者となるPV専門職自身に広く受け入れられるものとなるよう、設計、開発、導入前評価、および試験・パイロット運用・改訂の各段階に関与すべきです。

### 4.2.2. ヒューマン・オーバーサイトのメカニズム

ヒューマン・オーバーサイトは、異なる目的に対応し得るものであり、また各段階におけるガバナンス機構を通じて実現されます。<sup>6</sup>ヒューマン・オーバーサイトの具体的な方法は、監視対象となる活動の内容や、AIシステムに付与される自律性の程度に応じて多様に存在します。人間の介入の範囲、程度および強度に応じて、欧州委員会の「信頼できるAIのための倫理ガイドライン」では、三つの主要なガバナンス機構が示されています。すなわち、ヒューマン・イン・ザ・ループ（human-in-the-loop：HITL）、ヒューマン・オン・ザ・ループ（human-on-the-loop：HOTL）、およびヒューマン・イン・コマンド（human-in-command：HIC）です。HITLは、AIシステムの意思決定サイクルごとに人間が介入できる能力を指します。HOTLは、AIシステムにより高い自律性が与えられる場合を想定し、AIシステムの設計段階および運用監視の過程において人間が介入できる能力を指します。HOTLの特別な形態として「human on many loops」という概念があり、これは複数のAIモデルを同時に監視する際のスケーラビリティの課題に対応するものです。<sup>7</sup> HICは、AIシステムの全体的な活動、すなわち経済的、社会的、法的小および倫理的影響を含めて監督する能力、およびAIシステムをいつ、どのように使用するかを最終的に判断する能力を指します。

これには、特定の状況においてAIシステムを使用しないという判断、人間の裁量レベルの設定、またはシステムによる判断を上書きできる能力の確保が含まれます。<sup>6</sup> これら三つの用語の境界は参照資料によって異なる場合があります<sup>9</sup>、実際の運用方法も組織やユースケースに応じて異なります。例えば、PVプロセスを支援するAIシステムを構築する意思決定がなされた後（HIC）、人間による監督は、システムの使用文脈の定義や参照データセットの特定・作成を支援する開発段階（HOTL）から開始され得ます。さらに、導入後にはシステムの品質管理（QC）としての監督（HOTL）、あるいは半自動化システムにおける実行過程への直接関与（HITL）として実施される場合があります。

原則として、何らかの形の人間による監督は常に必要であり、主要または補助的なPVプロセスにおいてヒューマン・イン/オン・ザ・ループ（human-in/on-the-loop）が存在しない場合には、リスク評価によってその妥当性を裏付けるとともに、適切なリスク低減措置を講じる必要があります。

### 4.2.3 導入後AIシステムの監視および相互作用

AIシステムを監視し相互作用するために必要となる人間の介入の水準、頻度、方法および実施形態は、対象業務の複雑性、不適切な出力に伴うリスク、AIシステムの種類、およびその性能に依存します（第3章「リスクに基づくアプローチ」、第5章「妥当性と頑健性」参照）。AIに関する経験が蓄積されるにつれて、これらの要素の評価に関する明確性、指針および一貫性はさらに発展していくと考えられます。前述のとおり、特定のプロセスにおける人間とAIの役割分担は連続体として捉えることができ、AIが人間による評価および意思決定を支援するための準備作業のみを実施する段階から、人間が品質管理（QC）を実施しながら監視するほぼ完全自動化システムまで幅広く存在します。また、中間的な形態として、AIシステムが処理困難な症例を検出し、人間の専門家に提示するような運用も想定されます。

導入済みAIソリューションの性能を監視するために用いる指標および主要業績評価指標（KPI）は、試験およびバリデーション計画の一部として事前に定義しておく必要があります（第5章「妥当性と頑健性」、第3章「リスクに基づくアプローチ」参照）。

AIシステム単独での性能が十分でない場合（例えば、確立された人間の性能に達しない場合）、複雑または曖昧なケース、あるいは関連リスクが許容できないほど高い場合には、一つまたは複数の手作業プロセスを組み込み、最終的な出力について人間が完全に管理する必要があります。静的AIシステムがバリデーションにおいて人間の性能を上回った場合であっても、運用開始後に性能が許容水準を下回らないことを確認するため、継続的な監視が推奨されます（第5章「妥当性と頑健性」参照）。また、AIシステムに変更が加えられるたびに、ヒューマン・オーバーサイトは当該変更部分に重点を置き、変更対象に特化したサンプルを優先的に評価する必要があります。

導入後のAIソリューションの性能監視には複数の方法があります。静的AIシステムでは生成された出力の一部または全体を期待される結果と比較する事後的分析を実施することが可能です（第5章「妥当性と頑健性」参照）。その後、必要に応じて事後修正やモデルの再学習または再バリデーションを行います。より動的な運用として、意思決定支援環境において人間が独立した評価を実施し、AI出力を確認または修正するリアルタイムの相互作用も想定されます。このような動的AI運用では、相互作用自体がアルゴリズムへの即時フィードバックとなり、継続的な学習および調整を可能にします。また、主要AIシステムとは独立したモデルを並行稼働させる方法も、一時的または継続的な形で採用可能であり、これはAI支援型ヒューマン・オーバーサイト（AI-assisted human oversight）に該当します。

GenAI/LLMベースのシステムの監視には特に注意が必要です。ヒューマン・イン/オン・ザ・ループ (human-in/on-the-loop) に関与する人員は、出力の固有の変動性、説明可能性の限界、およびハルシネーションのリスクを理解し、AI結果に過度に依存しないよう留意する必要があります。運用プロセスは頑健であり、有効性が実証され、誤った出力が発生した場合でも信頼性を維持できなければなりません。特にハルシネーションは、一見整合的で説得力のある出力を生成し人間を誤導する可能性があります。AI技術の種類にかかわらず、PV専門家は自身の経験に基づきシステム出力に異議を唱えることができるよう権限を持つべきであり、自動化バイアスに陥らないよう注意する必要があります。一方で、確認バイアスの可能性にも留意し、予想外のAI出力であっても正しい可能性を排除しない姿勢が求められます。また、高性能なAIシステムにおいては、人間が極めて稀な誤りを見逃しやすいという「低有病率効果 (low prevalence effect)」が生じ得るため、特別な監視戦略が必要となる場合があります。<sup>8</sup>

### 4.3. 従来の役割の変革

PV領域がAI能力を取り入れ続けるにつれて、従来PV専門家が担ってきた一部の業務が置き換えられることにより、大規模なPV専門人材への依存度は低下すると予想されます。これにより一部の役割が時代遅れとなり、業務のオフショアリング時に職員に生じるものと同様に、PV人材の規模・多様性・経験が縮小する可能性があります。実際、PVプロセスにおける自動化およびAI利用の拡大により、PV専門家は反復的で時間を要する手作業から解放されます。その結果、特定の職務が不要となり、PV人材の規模、多様性、経験構成が縮小する可能性があります。これは、組織が業務をオフショア化した際に観察された人員構造の変化と類似しています。一方で、AI能力の急速な進展およびAI性能の動的な変化により、組織が将来の人員需要を正確に予測し、最適かつ持続可能な人的資源モデルを維持することは困難になる可能性があります。このような変化は、PV分野における雇用の将来性、職務喪失への懸念、さらには職場文化、モチベーション、充足感に関する不安を生じさせ得ます。また、一部のAIモデルが過去のデータセットやPV専門家による過去の判断記録を学習して構築されることから、不公平感が生じる可能性もあります。

一方で、AIの導入はPV専門家に新たな成長機会をもたらします。単純で時間を要する作業が減少することで、PV専門家はより科学的に高度で知的刺激の高い業務に集中できるようになります。さらに、AIシステムの導入に伴い、ガバナンスや人間による監督に関する新たな役割が生まれます。前述のとおり、PV専門家はAIモデルの試験、評価、実装監督および運用にますます関与するようになります。また、どの業務が自動化に適しているかを特定し、適切なAIユースケースを提案できる立場にあります。モデル学習およびバリデーションを含む設計・開発活動への参加、ユーザー受入試験への関与、既存プロセスの自動化および変更に伴う課題管理、監視および品質管理活動の実施、評価の不整合に関する問題の特定と解決、さらに自動化専門家やベンダーとの連携などが含まれます。

AIシステムの開発、利用および維持に関与することは、PV専門家が変化するPV環境に適応して進化する機会となりますが、そのためには従来のPV専門能力を超えたスキル拡張が求められます。<sup>9</sup>これには、新しいシステムの適切な利用能力、出力結果を批判的に評価する能力に加え、データサイエンスおよびAIに関する基礎的理解、すなわちAIの能力、リスクおよび限界に関する理解が含まれます。EU AI Actなどの規制枠組みでは、導入されたAIシステムを運用または使用する職員に対し、十分なAIリテラシーを確保することが組織の義務として求められています。<sup>6</sup>

PV専門家に限らず、QA部門の職員も、組織のヒューマン・オーバーサイト戦略およびAIシステムのバリデーション手法について理解を深め、ヒューマン・オーバーサイト活動が適切に実施されていることを確認する必要があります。同様に、PV向けAIソリューションの設計および開発に関与するAI専門家も、規制環境下で運用されるPVプロセスおよびその影響について理解を深める必要があります。

変更管理および組織準備戦略は組織の重要な責務であり、役割再定義およびスキル向上の取り組みにおいてPV職員を中心に据える必要があります。適切な変更管理および教育計画は、AIシステムをPVプロセスへ円滑かつ安全に統合し、職員の広範な関与と受容を促進し、各役割間の円滑な連携を実現するための前提条件となります（第9章「ガバナンスとアカウントビリティ」参照）。明確な学習体系およびキャリア発展フレームワークを伴う体系的な能力開発プログラムの導入も検討すべきです。

教育プログラムは、対象者（例：PVエンドユーザー、QA職員、AI専門家）の学習ニーズに適合するよう、内容および形式（教材および教育手法を含む）を慎重に設計し、文書化し、評価する必要があります。PV専門家、データサイエンティスト、QAチーム間のクロスファンクショナルトレーニングは有効なアプローチとなり得ます。意思決定支援環境における人材教育は、AIシステムを監視・運用する職員の教育手法として有用です。これは、特定ツールの操作方法を示すだけでなく、バックエンド機能の理解を促し自動化された出力を批判的に評価する能力を育成することを目的とします。例えば、既知の限界により人間による確認や判断が必要となる領域について職員を教育する必要があります。教育方法（集合研修またはオンライン、同期型または非同期型など）は、システムの複雑性、制約、対象ユースケース、意思決定ポイント、および組織または個人のニーズに応じて調整すべきです。<sup>10,11,12,13</sup>

## 第4章 – 参考文献

- 1 European Commission. Ethics guidelines for trustworthy AI. [Internet]. Brussels: European Commission; 2019 [Internet]. Brussels: European Commission; 2019. (Webpage accessed 21 March 2025)
- 2 World Health Organization (WHO). Ethics and governance of artificial intelligence for health: WHO guidance. Geneva: World Health Organization; 2021. (PDF accessed 19 September 2025)
- 3 European Parliament, Council of the European Union. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828. Off J Eur Union. 2024;L 1689:1-159. (Webpage accessed 21 March 2025)
- 4 Government of Canada. The Artificial Intelligence and Data Act (AIDA) – companion document. [Internet]. Ottawa: Government of Canada; 2022 (Webpage accessed 21 March 2025)
- 5 Ball R, Dal Pan G. “Artificial Intelligence” for Pharmacovigilance: Ready for Prime Time? Drug Saf. 2022; May;45(5):429-438. <https://doi.org/10.1007/s40264-022-01157-4> (Journal full text)
- 6 Enqvist L. ‘Human oversight’ in the EU artificial intelligence act: what, when and by whom? Law, Innovation and Technology. 2023;Jul3;15(2):508-535. <https://doi.org/10.1080/17579961.2023.2245683> (Journal full text)
- 7 Hillis JM, Payne K. Health AI needs meaningful human involvement: lessons from war. Nat Med. 2024;30:3397- 3398 <https://doi.org/10.1038/s41591-024-03311-0> (Journal abstract)

- 8 Rich AN, Kunar MA, Van Wert MJ, Hidalgo-Sotelo B, Horowitz TS, Wolfe JM. Why do we miss rare targets? Exploring the boundaries of the low prevalence effect. *J Vis.* 2008;Nov1;8(15):15. <https://doi.org/10.1167/8.15.15> ([Journal full text](#))
- 9 Danysz K, Cicirello S, Mingle E, Assuncao B, Tetarenko N, Mockute R, et al. Artificial Intelligence and the Future of the Drug Safety Professional. *Drug Saf.* 2019;Apr;42(4):491-497. <https://doi:10.1007/s40264-018-0746-z>. ([Journal full text](#))
- 10 Keen PG. Decision support systems: a research perspective. In: *Decision Support Systems: Issues and Challenges. Proceedings of the International Institute for Applied Systems Analysis (IIASA) Conference Series.* 1980;Jun23;11:23-27. (E-book abstract accessed 15 October 2025)
- 11 Wang RY. A product perspective on total data quality management. *Communications of the ACM.*1998;Feb1;41(2):58-65. <https://doi.org/10.1145/269012.269022> ([Journal full text](#))
- 12 Power DJ. A brief history of decision support systems. *DSSResources.com.* 2007;Mar10;3. ([Webpage](#) accessed 21 March 2025)
- 13 Turban E. *Decision support and business intelligence systems.* New Delhi: Pearson Education India; 2011.

## 第5章

# 妥当性と頑健性

### 原則

「妥当性」(validity)とは、システムが許容可能な範囲内で意図された目的を達成することを意味します。そのためには、許容される性能水準を事前に定義し、モデルの学習および／または試験に適切なデータを選択し、現実的な環境下でモデル性能を評価し、さらに継続的な品質保証(QA)プロセスへ統合することが求められます。

「頑健性」(robustness)とは、データのばらつきを考慮しながらも、システムが意図した目的を安定して達成できることを意味します。

### 主要メッセージ

- 医薬品安全性監視(PV)専門家および意思決定者は、AIシステムを自ら開発する場合であっても外部から導入する場合であっても、それらを批判的に評価できる能力を習得する必要があります。
- 想定される使用条件下で許容可能かつ頑健な結果を示す性能評価は極めて重要です。この評価は、AIモデルの目的を十分に検証できる幅広い事例を対象とし、多くの場合統計指標に基づいて実施されます。
- 関連するデータ型がテストセットに十分に含まれることを確保するよう注力することが重要です。これにより、バイアスの検出、想定される導入領域全体での適切かつ一般化可能な性能の促進、ユーザビリティの評価、およびモデル性能が低下する状況の特定が可能になります。
- 多くのPV用途では、まれな事象やパターン(例：安全性シグナル、重複症例)の識別が対象となるため、対象事象を強化したテストセットが必要となる場合があります。その場合、評価結果が実運用環境へ一般化可能であることを確保するため特別な注意が必要です。

## 5.1. はじめに

AIソリューションの妥当性と頑健性を確保することは、患者安全の確保、信頼性の構築そして利用者にとって最大限の価値を達成するうえで中核的な要素です。資源を最適に投入するためには、PV専門家および意思決定者は、AIシステムを自組織で開発する場合であっても外部から導入する場合であっても、提案されたAIシステムを批判的に評価できるようになる必要があります。そのためには、性能評価の基本原則および将来的な実運用性能に対する期待を誤らせる可能性のある一般的な落とし穴への理解が求められます。

AIモデルは多くの場合、PVユースケースを支援するより大きなコンピュータシステムの一部として組み込まれます。これらのシステムは、組織の標準的な手順に従ったコンピュータ化システムバリデーションの対象となるべきです。一般的には、これはAIコアモデル自体の妥当性と頑健性の確保とは別に扱われ、本書の対象範囲外となります。ただし、継続的に学習し新規データへ適応する動的AIモデルを含むシステムの検証に関する特有の考慮事項については、「継続的インテグレーションとデプロイ」の節で取り扱い

ます。本章では、AIモデルそのものの妥当性と頑健性確立に関する主要な考慮事項、すなわち学習および運用における基盤データへの依存性、および確率的・統計的性能評価の必要性に焦点を当てます。

PVデータの特性は、AI活用の可能性や適用方法に影響を与えます。AIモデルは学習データおよび予測に用いるデータの品質に強く依存しますが、PVデータには不整合、不完全入力、不正確性が含まれることが多く、情報源によって大きく異なります。例えば、有害事象報告の多くは自発報告であり、報告慣行は時期、組織、報告者の種類によって変化します。これは報告される有害事象の種類、取得される情報、およびコード化方法に影響を及ぼします。このような不整合や不正確性はモデル精度の低下を招き体系的なばらつきはAIモデルの一般化可能性を低下させ、データドリフトへの感受性を高める可能性があります。また、地域や組織間で一貫した性能を維持することを困難にする場合があります。

一般に、有害事象報告の品質と一貫性のばらつき、および薬剤—事象関係の複雑性により、PV領域ではAIソリューションの妥当性と頑健性を確保するため、他分野よりも広範な人間の関与が必要となる場合があります。PV業務は規制対象であり、妥当性と頑健性に関する規制当局の期待は、企業内部の期待とは異なる場合があります。

性能評価および試験は、AIモデルの妥当性と頑健性確保における重要な要素であり、本章の中心的テーマです。これらの考慮事項は、AIモデルの開発および学習段階においても同様に反映されることが望ましいとされています。例えば、機械学習分類器<sup>1</sup>におけるバイアス低減は性能改善につながる場合があります、異なる誤りに異なるコストが伴う場合にはコスト感度学習も有効となり得ます。<sup>23</sup> 一方で、高性能を達成するために必ずしもそれらが必要または実行可能とは限りません。例えば、大規模言語モデルは特定訓練を受けていない言語課題に対してもゼロショット学習で良好な性能を示す場合があります。また、PV組織が既に開発済みのAIシステムを導入する場合、モデル開発に関与できないこともあります。いずれの場合でも、想定使用条件を反映した独立試験において適切な性能が実証されることを確認することが重要です。

## 5.2. 仕様・設計

### 5.2.1. ユースケースおよび適用領域

PVにおけるAIソリューションの想定ユースケースおよび適用領域は明確に定義されるべきであり、可能な限りそれらに対応した性能評価が実施される必要があります。例えばPVシグナル検出手法の評価では、既知の副作用よりも過去の安全性シグナルの方が評価基盤として適切である場合が多く、これは報告パターンが重要な点で異なるためです。<sup>4</sup> 同様に、自由記述テキストから有害事象を認識するAIモデルを広範囲用途で使用する場合には、様々な医薬品、有害事象、患者および医療従事者からの報告、関連言語などを含む評価が必要となります。理想的には、学習およびテストデータには関連するデータ型が十分に含まれており、適用領域全体において一般化可能な性能を確保し、ユーザビリティ評価、バイアス検出、および性能低下が生じる状況の特定を可能にする必要があります。より高度なAIシステムでは、実行可能なタスクと拒否すべきタスクを定義する拒否ポリシーの設定も検討対象となります。

AIシステム設計では、過学習への感受性、計算複雑性、外れ値に対する頑健性も考慮されるべきです。特にテストセットが十分に大規模または多様でない場合、これらの影響を性能評価で十分に把握できない可能性があります。高度な専門知識に依存する複雑な手法は、同等の専門性が存在しない環境では他用途へ容易に移転できない場合があります。実運用では、理論的優位性よりも、特定目的に対して実際に最良の性能を示す方法が重視されます。

## 5.2.2. 学際的連携

AIモデルの妥当性と頑健性を確保するためには、多分野にわたる協働が必要となることが多く、PV意思決定者および実務担当者に加え、データサイエンティスト、AI専門家、コンピュータシステムバリデーション経験者などの関与が求められます。多様な視点と専門知識、PVシステムへの統合方法に関する深い理解、および期待される利益と関連リスクの明確化は、AIソリューションがライフサイクル全体にわたり有効に機能し、特定されたニーズに対応することを確実にします。

薬剤と有害事象の複雑な関係を扱うAIシステムでは、基盤となるPVデータの品質や由来のばらつきを考慮すると、特にヒューマン・イン・ザ・ループ（Human-in-the-Loop：HITL）が必要となる場合が多くあります。このような用途におけるAI出力は、臨床的背景、既知の薬理学的機序、およびデータ上に存在しない可能性のある代替説明を踏まえて解釈される必要があり、<sup>5,6</sup> 人間の介入により最終結果の臨床的妥当性および科学的妥当性が確保されます。一方で、個人情報やマスキングや医薬品・有害事象コード付与のような比較的単純な業務は、人間の介入を最小限とした自動化に適している場合があります。

## 5.2.3. 参照基準の定義

テストセットは、想定される適用領域と整合し、現実的条件下での性能を示せるものでなければなりません。想定する用途に関連する参照基準は明確に定義され、最新状態に維持される必要があります。多くのPV用途では、対象業務を人間が実施した結果が参照基準となり、実例データを専門家が分類（アノテーション）する形で構築されます。この際、複数評価者による注釈付与などを通じて不一致を低減する方法がしばしば必要となります。過去の人間による注釈を利用する場合には、分類定義を遡及的に明確化し、すべての注釈が統一基準に適合し将来用途に適切であることを確認する必要があります。旧来の基準や異なるデータ種別に基づく注釈は除外が必要となる場合があります。新規に参照基準を構築する場合には、明確な注釈ガイドラインの策定が推奨されます。これは既存PV業務手順の明確化や強化につながり、暗黙的ではらつきのあった判断プロセスを組織内で統一する契機となる場合があります。プロジェクト規模によっては、性能評価用テストセットの注釈がAIモデル開発から独立して実施されることを確保する必要があり、例えば特定モデルを知らない評価者による注釈が望ましいとされます。同様に、人間とAIのチーム評価を行う場合には、人間側の資格要件が想定ユースケースと適用領域に一致している必要があります。

場合によっては参照基準の分類境界が明確でなく、解釈の曖昧さが生じることがあります。例えば、2件の有害事象報告を同一事象とみなす確信度の基準は組織ごとに異なる場合があります。用途によっても判断基準が変わることがあります。人間が最終確認する場合には疑いのある重複を広く抽出できる一方、自動削除を行う場合にはより慎重な基準が必要となります。同様の曖昧さは、自由記述をMedDRA等の標準用語へ対応付けるNLPタスクでも生じます。また、シグナル検出や因果関係評価では、専門家間でも判断が一致しないことが珍しくありません。

PV分野では、参照データセットへの持続的かつ再利用可能なアクセス確保が長年の課題となってきました。AIソリューションの広範な影響可能性を踏まえると、参照基準の作成方法および前提条件を明確にしたうえで、最新かつ利用可能な状態で維持することが重要です。

### 5.3. 性能評価

AIモデルを適切に評価するためには、性能評価が不可欠です。性能評価を実施または評価できる能力は、AIシステムを開発する者にとっても、AIシステムの導入提案を受ける者にとっても重要な技能です。

PVにおけるAIモデルの性能評価に関連する多くの指標は情報検索分野に由来しており、主として二値分類問題として捉えられるユースケースに適用されます。二値分類では、手法によって検出されることが望ましい事例を陽性（positive controls）、検出されるべきでない事例を陰性（negative controls）と呼びます。本節ではこの用語を用います（場合によっては陽性を対象事象と表現します）が、順位付け評価、教師なし学習、コンテンツ生成など、異なる評価枠組みを必要とするユースケースが存在することも認識されています。

AIシステムは人間と同様に、より複雑な分類課題において完全な性能を達成することは通常ありません。実際の運用環境では、専門家であっても正しい分類が本質的に曖昧となる事例が存在します（例えば情報不足の場合）。そのため、性能は通常、テストセットと呼ばれる症例サンプルを用いて統計的に評価されます。再現率（Recall）は、AIソリューションが対象事象をどれだけ正しく検出できたかを示し、感度（Sensitivity）と同義語です。適合率（Precision）は、AIが抽出した事象のうち対象事象が占める割合を示し、陽性的中率（Positive Predictive Value：PPV）と同義語です。

適合率と再現率（同様に感度と特異度）のバランスは通常調整可能であり、異なる種類の誤りに伴うコストや正しい判断の有用性に基づいて決定されるべきです。F1スコアのような複合指標（適合率と再現率の調和平均）は、一定の前提条件のもとで両者を同時に考慮した単一指標として予測精度を表します（F1スコアでは適合率と再現率が同等に重要であり、偽陽性と偽陰性のコストが同等であると仮定します）。テストセットは、想定される適用領域を十分に反映し、統計的に信頼できる性能推定を可能にするため、十分な規模、多様性、および代表性を備えている必要があります。また、異なる集団を含み、想定用途に沿った様々なシナリオを考慮する必要があります。<sup>1</sup>

主な関心は将来的な実運用におけるAIソリューションの性能であるため、性能評価は開発時に直接使用されたデータから独立して実施される必要があります（開発段階でのクロスバリデーションや学習・検証データ分割とは別の要件です）。利用者主導の設計判断は、開発者がテストセットへ初めてアクセスする前に確定していなければなりません。これは、モデル構造、ハイパーパラメータ、初期化方法など多数の分析選択肢を持つ複雑な手法において特に重要です。<sup>7</sup>開発と評価の間に生じ得る依存関係は検討および排除される必要があり、最も明白な例は同一データ点が両段階で使用される場合です。さらに微妙な依存関係も存在し、過度に楽観的な性能推定につながる可能性があります。例えば、学習セットとテストセットが同一の医薬品や有害事象の部分集合を共有している一方で、実際の適用領域がより広範である場合などが挙げられます。<sup>8</sup>

### 5.3.1. ベンチマーキング

理想的には、利用可能な場合、性能は関連するベンチマーク手法と比較して評価されるべきです。例えば、AIに基づくシグナル検出手法は、当該組織における基準手法が標準的な不均衡解析（disproportionality measures）である場合、これらと比較されることがあります。AIモデルに基づくものを含む、より複雑なベンチマーク手法を用いる場合には比較対象として適切に機能するよう、対象タスクに合わせてベンチマーク手法が適切に実装および調整されていることを慎重に確認する必要があります。

公開されたベンチマーク用テストセットが存在する場合、それらを用いた性能評価を実施できますが、これは対象とする導入領域に即した性能評価を補完するものとして用いることが望まれます。現時点では、PVにおける公開ベンチマークは一部の特定用途に限られています。これには、新規安全性シグナルのセット、<sup>9,10</sup> 既知の副作用反応のセット<sup>11,12,13,14,15</sup> ならびに臨床的に重要な薬物間相互作用のデータセットが含まれます。<sup>16</sup> しかしながら、これらベンチマーク参照セットへの継続的なアクセスを維持することは課題となる場合があります、維持管理および更新の程度にもばらつきがあります。

全体的な性能評価を補完する目的で、サブグループ解析を実施することで、導入領域の異なる部分におけるAIモデルの強みおよび弱みについて有用な情報が得られます（第8章「公正性と衡平性」参照）。同様に、感度分析は、仕様や設計の変動に対するAIモデルおよびその評価結果の頑健性を評価するうえで有用です。

<sup>i</sup> For a continually updated inventory, see for example <https://oecd.ai/en/catalogue/metrics>

### 5.3.2. 低頻度事象環境における特別な考慮事項

多くのPV業務では、まれなパターンや事象の認識に重点が置かれます。例えば症例検索タスクでは、妊娠、投薬過誤、再投与陽性（positive rechallenge）、薬物性肝障害など特定のテーマに対して、ほとんどの報告は通常関連性を持ちません。同様にPVにおけるシグナル検出では、ほとんどの薬剤-事象の組合せは真の副作用反応ではなく、まして新規に検出される安全性シグナルであることはまれです。これらの希少事象を効果的に管理・解析するためには信頼性の高い参照データセットが必要ですが、SIDERのような既存資源は情報が古く静的である場合が多く、代替的な解決策の必要性が示唆されています。<sup>17</sup> さらに極端な例として、大規模な個別症例報告データベースにおける重複症例の組合せは極めてまれです。例えば100万件のデータベースにおいて10%の報告が1件の重複を持つ場合でも、無作為に選択した2報告が重複である確率は1,000万分の1にすぎません。<sup>ii</sup>

このような陽性例の低頻度（クラス不均衡）は、正確な性能評価を困難にするため、特別な注意と配慮が必要です。例えば、アノテーションの品質向上のために複数評価者による二重アノテーションを行うかどうかなど、各アノテーションの品質とテストセット規模（あるいは作成に要する時間・コスト）との間でバランスを取る必要があります。また、単純な無作為抽出によるテストセットでは陽性例が不足する一方、陽性例を意図的に増やしたテストセットでは適合率（precision）や再現率（recall）の推定値が誤解を招く可能性があります。この点についてはNorénら（2025）に詳述されています。<sup>18</sup>

ヒューリスティック手法によってテストセット中の対象事象割合を増加させた場合、AIモデルにとって識別が困難な対象事象が含まれにくくなる可能性があるため、再現率が過大評価されることがあります。これは再均衡化手法を避けるべきことを意味するものではありませんが、使用する場合にはその影響を認識し、批判的に評価する必要があります。

同様に、適合率はテストセットにおける対象事象の有病率に強く依存します。対象事象を増強したテストセットでは基礎有病率が人工的に高くなるため、単純な適合率推定は楽観的になります。信頼性のある適合率推定のためには、テストセット中の陽性例の割合を、可能な限り実際の導入環境に一致させる必要があります。特定のAIモデルに対しては、無作為抽出したサンプルにモデルを適用し、抽出されたすべての事例をアノテーションすることで適合率を比較的容易に推定できます。ただし、この方法で作成されたテストセットは当該AIモデルに依存するため、モデルが変更された場合には再作成または拡張が必要となり、再現率評価には適していません。

適合率および再現率の推定値は、選択された意思決定閾値に依存します。そのため性能評価は、想定される導入環境において偽陽性と偽陰性のバランスが適切となる意思決定閾値を前提として実施する必要があります。

<sup>ii</sup>  $0.10 \times 1/10^6$

### 5.3.3. 要約統計を超えて

前節で述べた指標により示される要約統計は、AIモデルの性能を評価および理解するうえで有用ではありますが、それだけでは十分ではありません。テストセットにおける個々の事例に対するAIモデルの分類結果について、代表的で具体的な例にアクセスし、それを詳細に確認できることも重要です。エラー解析の段階で偽陽性および偽陰性を検討することは、それぞれAIソリューションおよびその評価の強みと限界に関する有益なインサイトを与えます。例えば、匿名化処理における偽陰性が「Mr」に続く氏名の未削除であった場合、全体として再現率が優れていたとしても、その誤りが単純に見えるため、エンドユーザーの信頼を損なう可能性があります。一方、偽陰性が「AF」であり、それがイニシャルなのか心房細動 (atrial fibrillation) の略語なのかを周囲の文脈から専門家であっても判断が難しい場合には、全体の適合率指標はむしろ保守的であると考えるべきかもしれません。正しく分類された事例のレビューもまた、AIシステムが困難な課題をどの程度解決できているかについてのインサイトを与えます。すなわち、難易度の高い症例も正しく分類できているのか、それとも容易な症例のみを正しく処理しているのかを確認する必要があります。これは、比較対象となる基準手法が存在しない場合、全体的な性能指標だけでは課題の難易度を理解できない可能性があるため、特に重要です。比較対象となる基準手法が存在する場合には、両手法で異なる分類結果となった事例をレビューすることで、提案されたソリューションが比較手法に対してどのような点で性能向上を示しているのかをより深く理解することができます。

### 5.3.4. 教師なし学習

クラスタ解析などの教師なし学習を実行するAIシステムでは、人間の手で注釈付けされた参照セットを用いずに、データ駆動型の方法でパターンが識別されます。そのため、性能評価には別のアプローチが必要となります。場合によっては、AIの出力結果について人間による主観的なレビューおよび評価に依拠することがありますが、その際には認知バイアスの可能性を考慮し、これを軽減する必要があります。ひとつの方法として、複数の異なるAIモデルの結果をブラインド化したうえで領域専門家に提示し、どの結果をより適切と判断するかを評価してもらう方法が考えられます。また、教師なし学習のために特別に設計された性能評価手法も存在します。例えば、侵入者検出分析 (intruder detection analysis) では、AIモデルが関連すると判断してグループ化した項目群の中に無関係な「侵入者」を混在させ、専門家にそれを特定してもらうことで、一貫性を侵入者検出率によって評価します。<sup>19</sup>

### 5.3.5. 生成型出力

生成AI (GenAI) モデルは、あらかじめ定義された選択肢 (例: Yes/NoやMedDRA Preferred Term) への限定や追加的な後処理を行うことなく利用可能な、自由形式で多くの場合より長文のテキスト (またはその他のコンテンツ) を生成できます。このような応用例には、文章要約、翻訳、報告書作成、一般向け表現への書き換えなどが含まれます。通常、この種のタスクには唯一の正解出力は存在せず、文章の流暢性や一貫性といった側面を、タスク固有の評価指標とともに評価する必要があります。本分野は急速に進展しており、本書執筆時点では複数の評価指標を並行して用いることが一般的です。例えば、可読性や有害性に関する汎用指標を算出できるほか、正解となる参照テキスト (例: 翻訳や要約タスク) が存在する場合には、構文的重複度 (例: BLEU: Bilingual Evaluation Understudy) や意味的重複度 (例: BERT score) を計算できます。同様に、主要エンティティの保持状況についても、人間の手で注釈付けされた参照セットが存在すればそれに基づき評価できます。人間による評価を前向きに実施することも可能ですが、

その場合は前述の教師なし学習の文脈と同様に慎重な設計および実施が必要です。例えば、ペアワイズ選好評価やAI出力に支援された人間のタスク遂行成功率の評価などが選択肢となります。人間による評価は依然としてゴールドスタンダードですが、包括的な性能評価の一環として、生成型大規模言語モデル（LLM）自体にAIシステム出力を複数の観点から評価させる方法もあります。このような「LLM-as-a-judge」アプローチは拡張性に優れていますが、評価対象モデルとは独立した高性能な評価モデルが必要であり、通常は人間によるキャリブレーションが求められます。PV領域におけるGenAI応用の性能評価の最近の例としては、有害事象（AE）報告の要約<sup>20</sup>や、COVID-19ワクチン報告における個別症例の因果関係評価に関するLLM生成臨床推論の評価が挙げられます。<sup>21</sup>

### 5.3.6. 非決定論的システム

決定論的AIシステムは、同一の入力に対して常に同一の出力を生成します。サポートベクターマシンや決定木（decision tree）などの予測モデルはこの性質を有します。また、特定の大型言語モデル（LLM）（例：BERTのようなマスク型エンコーダ）や分類タスクに用いられるその他のディープニューラルネットワークも、学習またはファインチューニング完了後に重みが固定されれば同様です。これは、モデル学習過程に確率的要素が含まれており、同一データで再学習した場合に異なるパラメータが得られる可能性がある場合でも当てはまります。

これに対し、クラスタ解析、ネットワーク解析、有害事象（AE）の意味ベクトル表現をデータ駆動的に導出する手法などの教師なし学習は確率的性質を持つ場合があり、同一データに対して繰り返し実行しても異なる結果を生成することがあります。同様に、生成型LLMも、基盤モデルに変更がなくても、同一プロンプトに対して異なる出力を生成するのが一般的です。このようなAIモデルでは、出力の安定性、すなわち繰り返し解析した際の結果の類似度が重要な追加的性能指標となる場合があります。場合によっては確率性の程度をハイパーパラメータによって直接制御できることもあります。また、手法によっては出力の変動性を低減するための特別な対策が講じられることもあります。疑似乱数生成器のシード固定により結果の再現性を人工的に確保できる場合もありますが、これはプロプライエタリモデルでは不可能なことがあり、またAIソリューション固有の安定性そのものを改善するものではないため、別途評価が必要です。非決定論的AIシステムが特定の用途に適切かどうかは、その使用文脈および出力変動がもたらし得る負の影響を踏まえ、リスクに基づくアプローチの原則に基づいて判断する必要があります。

## 5.4. ヒューマン・イン・ザ・ループによる人工知能システムの評価

多くのAIシステムはインテリジェンス・オーグメンテーション、すなわち人間の意思決定を支援強化することを目的としています。この文脈では、性能評価の主な対象は人間とAIから成るチーム全体となり、前述とは異なる性質の試験設計が求められます。現時点では、PV領域におけるこの種の研究の経験は限定的ですが、少なくともチームを構成する人間のメンバー間の技能や選好のばらつきを考慮する必要があります。また、適切なテストセットの定義にも新たな課題が生じ得ます。例えばシグナル検出では、人間の専門家過去の安全性シグナルから盲検化することは困難であり、さらに人間単独の専門家による分類品質を上回ることを目標とする場合、参照基準を設定すること自体が難しい場合があります。

許容可能な性能の定義は、AIシステムがPVシステムへどのように統合され、人間がどのように当該システムと相互作用するかを踏まえて検討する必要があります。<sup>22</sup> 例えば、自然言語処理（NLP）を用いて原資料から有害事象（AE）を識別・抽出するシステムの性能評価では、全体的な性能指標に加え、結果中の誤りが容易に発見できるか、また人間とAIを組み合わせたエンドツーエンドのハイブリッドプロセスが完全手作業と比較して優れているかどうかを評価対象に含めることが考えられます（例：Parkら 2023<sup>23</sup>）。

## 5.5. 継続的インテグレーションとデプロイ

本番環境に導入されたAIモデルは、性能が維持または向上していることに重点を置き、実運用下で継続的に監視される必要があります。状況によっては、業務側の理解の深化や外的要因による業務条件の変化に応じて、本番運用中に性能基準を見直し、更新する必要が生じる場合があります。

機械学習（ML）コンポーネントを含む導入済みAIソリューションについては、性能劣化やデータセットドリフトによる悪影響のリスクを管理するため、定期的な再学習に関する適切なプロセスおよび品質管理（QC）が求められます。再学習は、既存のモデルアーキテクチャ内での段階的なファインチューニングにとどまる場合もありますが、導入領域の大きな変更がある場合には、AIモデルのアーキテクチャ自体の変更が必要となることがあります。後者は、医薬品からワクチンへの対象範囲変更、基盤となる医療用語体系やデータ構造の改訂、規制または慣行の更新などにより生じ得ます。

継続的な性能評価は、AIシステムがMLコンポーネントを含むか否かにかかわらず重要です。その頻度はリスクに基づくアプローチに従って決定され、重大なデータドリフトや性能低下を検出するためのデータ駆動型の安全措置を含む場合があります。このような兆候が確認された場合には、追加評価、再学習の実施、アルゴリズム使用の停止、または結果への信頼性を維持するためのQC強化などの是正措置が必要となる場合があります。導入後に（準）リアルタイムで継続的に更新される動的AIモデルでは、モデルドリフトの自動検出が再バリデーション活動の開始契機となることもあります。このような状況でAIソリューションを再導入する際には、関連活動および受入基準の文書化が求められる場合があります。例えば、AIシステム変更時に確認する既知の入力／出力ペアの設定や、自動化バイアスを防止する仕組みなどが含まれます。

MLの潜在的利点の一つは、リアルワールドデータ（RWD）からの学習を含む反復的改良により性能を向上できる点にあります。この考え方を支援する枠組みとして、米国FDA、カナダ保健省（Health Canada）、および英国医薬品・医療製品規制庁（MHRA）は、ML対応医療機器ソフトウェア機能（ML-DSF）に対する「事前規定変更管理計画（Predetermined Change Control Plan）」を提示しています。その一般原則は、PV領域のAIシステムにも適用可能と考えられます。事前規定変更管理計画には通常、1)計画された具体的変更内容の詳細説明、2)アルゴリズムの性能が継続的に許容範囲内に維持されることを保証するための開発・検証・実装方法論、3)変更による便益およびリスクとその低減策に関する影響評価が含まれます。変更内容の詳細説明には、変更実施によって生じるアルゴリズム特性および性能の変化を含める必要があります。変更の例としてはMLモデルの再学習が挙げられます。変更の開発・評価・実装に用いるデータおよび方法を詳細に定義したプロトコルを作成し、これに従う必要があります。また、変更に関する影響評価を実施し、特定されたリスクを管理するための低減策を策定する必要があります。このアプローチは、変更対象となるPVプロセスを統括する品質マネジメントシステム（QMS）へ組み込まれるべきです。

さらに、発生した問題や異常を容易に報告できる仕組みを整備し、それらに迅速に対応し、必要に応じて適切な水準へとエスカレーションする必要があります。理想的には、フィードバック受領の確認、調査状況の共有、必要なAIシステム変更の実施までを含む対応体制が求められます。

## 第5章 – 参考文献

- Hort M, Chen Z, Zhang JM, Harman M, Sarro F. Bias mitigation for machine learning classifiers: a comprehensive survey. *ACM J Responsib Comput*. 2024;1(2):Article 11. <https://doi.org/10.1145/3631326> (Journal full text)
- Elkan C. The foundations of cost-sensitive learning. In: *Proceedings of the 17th International Joint Conference on Artificial Intelligence (IJCAI'01)*, Vol. 2. San Francisco (CA): Morgan Kaufmann Publishers; 2001;973-978. (Journal full text)
- Araf I, Idri A, Chairi I. Cost-sensitive learning for imbalanced medical data: a review. *Artif Intell Rev*. 2024;57:80. <https://doi.org/10.1007/s10462-023-10652-8> (Journal full text)
- Norén GN, Caster O, Juhlin K, Lindquist M. Zoo or savannah? Choice of training ground for evidence-based pharmacovigilance. *Drug Saf*. 2014;Sep;37:655-659. <https://doi.org/10.1007/s40264-014-0198-z> (Journal abstract)
- Bradford Hill A. The environment and disease: association or causation? *Proc R Soc Med*. 1965;58:295-300. <https://doi.org/10.1177/003591576505800503> (Journal full text)
- Meyboom RHB, Hekster YA, Egberts ACG, et al. Causal or casual? *Drug Saf*. 1997;17:374-389 <https://doi.org/10.2165/00002018-199717060-00004> (Journal full text)
- Duin RP. A note on comparing classifiers. *Pattern Recognit Lett*. 1996;May 1;17(5):529-536. [https://doi.org/10.1016/0167-8655\(95\)00113-1](https://doi.org/10.1016/0167-8655(95)00113-1) (Journal abstract)
- Rudin C, Radin J. Why Are We Using Black Box Models in AI When We Don't Need To? A Lesson from an Explainable AI Competition. *Harv Data Sci Rev*. 2019;1;1-9. <https://doi.org/10.1162/99608f92.5a8a3a3d> (Journal full text)
- Harpaz R, Odgers D, Gaskin G, DuMouchel W, Winnenburg R, Bodenreider O, Ripple A, Szarfman A, Sorbello A, Horvitz E, White RW. A time-indexed reference standard of adverse drug reactions. *Scient Data*. 2014;Nov11;1(1):1-0. <https://doi.org/10.1038/sdata.2014.43> (Journal full text)
- Sartori D, Aronson JK, Norén GN, Onakpoya IJ. Signals of adverse drug reactions communicated by pharmacovigilance stakeholders: a scoping review of the global literature. *Drug Saf*. 2023;Feb;46(2):109-120. <https://doi.org/10.1007/s40264-022-01258-0> (Journal full text)
- Coloma PM, Avillach P, Salvo F, Schuemie MJ, Ferrajolo C, Pariente A, et al. A reference standard for evaluation of methods for drug safety signal detection using electronic healthcare record databases. *Drug Saf*. 2013;Jan;36(1):13-23. <https://doi.org/10.1007/s40264-012-0002-x> (Journal abstract)
- Ryan PB, Schuemie MJ, Welebob E, Duke J, Valentine S, Hartzema AG. Defining a reference set to support methodological research in drug safety. *Drug Saf*. 2013;Oct;36(Suppl 1):S33-S47. <https://doi.org/10.1007/s40264-013-0097-8> (Journal abstract)
- Kuhn M, Letunic I, Jensen LJ, Bork P. The SIDER database of drugs and side effects. *Nucleic Acids Research*. 2016;Jan4;44(D1):D1075-1079. <https://doi.org/10.1093/nar/gkv1075> (Journal full text)
- Demner-Fushman D, Shooshan SE, Rodriguez L, Aronson AR, Lang F, Rogers W, Roberts K, Tonning J. A dataset of 200 structured product labels annotated for adverse drug reactions. *Scientific Data*. 2018;Jan30;5(1):1-8. <https://doi.org/10.1038/sdata.2018.1> (Journal full text)
- Bayer S, Clark C, Dang O, Aberdeen J, Brajovic S, Swank K, Hirschman L, Ball R. ADE eval: an evaluation of text processing systems for adverse event extraction from drug labels for pharmacovigilance. *Drug Saf*. 2021;Jan;44:83-94. <https://doi.org/10.1007/s40264-020-00996-3> (Journal full text)
- Kontsioti E, Maskell S, Dutta B, et al. A reference set of clinically relevant adverse drug-drug interactions. *Sci Data*. 2022;9:72. <https://doi.org/10.1038/s41597-022-01159-y> (Journal full text)
- Painter J, Powell G, Bate A. PVLens: enhancing pharmacovigilance through automated label extraction. [preprint].arXiv. 2025;Mar26. <https://doi.org/10.48550/arXiv.2503.20639> (Journal full text)

- 18 Noren N, Meldau E-L, Ellenius J. Critical appraisal of artificial intelligence for rare-event recognition: principles and pharmacovigilance case studies. arXiv:2510.04341 [cs.LG] <https://doi.org/10.48550/arXiv.2510.04341> (Journal full text)
- 19 Chang J, Boyd-Graber J, Gerrish S, Wang C, et al. Reading tea leaves: how humans interpret topic models.[unpublished manuscript]. 2009. (Full text accessed 27 April 2025)
- 20 Dowdy K, Hoffman A, Giles T, Kugele D, et al. Summarizing FAERS Narratives with Generative AI: Methods, Resource Requirements, and Quality Assessment. FDA, SYMPOSIUM, SCIENTIFIC COMPUTING + DIGITAL TRANSFORMATION, 2024. Center for Drug Evaluation and Research, U.S. Administration, Booz Allen Hamilton. (PDF accessed 23 September 2025)
- 21 Abate A, Poncato E, Barbieri MA, et al. Off-the-shelf large language models for causality assessment of individual case safety reports: a proof-of-concept with COVID-19 vaccines. Drug Saf. 2025;48:805-820. <https://doi.org/10.1007/s40264-025-01531-y> (Journal full text)
- 22 The US Food and Drug Administration (FDA). Good machine learning practice for medical device development: guiding principles. The US Food and Drug Administration: Silver Spring, MD, USA, 2021. (Webpage accessed 21 March 2025)
- 23 Park J, Djelassi M, Chima D, Hernandez R, Poroshin V, Iliescu AM, Domalik D, Southall N. Validation of a natural language machine learning model for safety literature surveillance. Drug Saf. 2024;Jan;47(1):71-80 <https://doi.org/10.1007/s40264-023-01367-4> (Journal abstract)

## 第6章 透明性

### 原則

人工知能（AI）に関する透明性とは、組織間または個人間において情報を開示することを指します。これには、AIシステムのライフサイクル（すなわち、設計、開発、評価、導入、運用、再学習、保守および廃止）に関する関連文書を共有し、トレーサビリティを確保することが含まれます。また、ステークホルダーがAIシステム、その利用方法、リスク、限界、想定される利点、および自身の権利への影響について全体像を理解できるよう、十分な情報を提供することが求められます。

### 主要メッセージ

- コアとなる医薬品安全性監視（PV）業務において、いつ、どのようにAIシステムが使用されているかを明示することは、専門家、意思決定者、規制当局および一般社会からの信頼を構築するうえで極めて重要です。
- コアPV業務に導入されるAIソリューションの性質については、モデルアーキテクチャ、想定される入力および出力、ならびに人間とコンピュータの相互作用のレベルおよび形態を含め、十分に説明される必要があります。
- PV用途におけるAIモデルの有効性および限界を明確に示すためには、当該タスクに対する性能評価結果を提示し、テストセットの範囲および性質（参照基準の定義およびサンプリング戦略を含む）を説明する必要があります。
- 提示される性能指標は、想定される導入領域に適合したものであり、関連するベンチマークとの比較を行うとともに、正しい出力および誤った出力の代表例に対する定性的レビューによって補完される必要があります。
- 可能な場合には、AIモデルがどのような一般原理および論理に基づいて機能し、結果または予測に到達するのかについて説明を共有することが望まれます。説明可能性が限定的である場合には、その点を明確に認識し、議論する必要があります。

## 6.1. はじめに

透明性は、AIシステムの性質および利用方法に関して、ステークホルダーへ関連情報を提供することを目的とします。これは、AIシステムを開発または導入する主体が、主要なステークホルダーに対してどのような情報を共有するかを示す概念です。透明性の主な目的は、信頼の構築、開発に関与していない個人や組織がAIシステムの設計および性能を検証・精査できるようにすること、ならびに規制要件への適合を確保することにあります。

ガバナンスとアカウントビリティに関する章でさらに詳述されているように、AIシステムのライフサイクルの各段階において、透明性および情報開示の主たる方向性は異なります。例えば、設計段階では事業責任者がAIシステムの仕様および要求事項について開発者に対して透明性を確保する必要があります。一方、導入前段階では、開発者がAIシステムの性質および性能について事業責任者に対して透明性を確保する必要があります。

日常運用段階においては、組織からエンドユーザー（場合によっては規制当局を含む）への情報開示が最も重要な透明性の形態となる場合があります。

## 6.2. 人工知能利用の開示

PVの各業務において、なぜ、いつ、どのようにAIが使用されているのかを開示することは不可欠です。これは、開発者、PV専門家および意思決定者、規制当局、医療従事者（HCP）、患者を含むステークホルダーの間における信頼、認識および責任を維持するためです。予測または出力に対するAIモデルの確実性を示す信頼度スコアやその他の指標は、このような開示の有用な要素となり得ます。ただし、これらのスコアおよび指標の妥当性と頑健性も確保されなければならない、その意味がエンドユーザーに対して明確に説明される必要があります。

規制当局は、適用される法令および規制への適合を保証するため、AI利用の開示を求めています。この目的のため、ソフトウェアベンダーおよび社内開発部門はPV組織に対して透明性を確保する必要がある、PV組織はさらに規制当局に対して透明性を確保する必要があります。同時に、PVデータの処理または解析にAIソリューションを使用する個人は、自らの業務プロセスにおけるAIの役割について情報提供を受ける必要があります。これにより、AIを十分な理解のもとで業務に統合し、その有効な活用を支援するとともに、AI利用に起因する問題を特定できるようにします。

また、PV専門家は、データ要素の由来およびその取得または作成にAIソリューションが関与したかどうかについても説明する必要があります。PVデータの人間による解釈は、その取得方法に依存する場合があります。例えば、安全性シグナル評価者は、構造化データから自動生成された症例記述と、患者または医療従事者による逐語的記述に基づく症例記述とは、異なる重み付けを行う可能性があります。また、その由来が適切に開示されていない場合、AI生成情報が後続のAIモデル開発における参照基準の一部として再利用されるという悪循環が生じるリスクも存在します。

## 6.3. 人工知能モデルに関する透明性

PVにおいて使用されるAIモデルの十分な透明性を確保することは、信頼の醸成、十分な情報に基づく意思決定の促進、およびこれらのモデルが適切に適用されることを保証するうえで極めて重要です。理想的には、透明性はAIモデルに基づいてPV専門家が行った意思決定についても捕捉するまで拡張されるべきです。モデル透明性は倫理的な要請でもあり、すべての関係者が使用しているシステムを理解し、その出力に基づいて十分な情報に基づいた意思決定を行えることを確保します。以下に、ステークホルダーに開示すべきAIモデルの主要な側面を示します。また、モデルが意図された用途およびステークホルダーのニーズと整合していることを確保するため、設計上の選択の根拠についても説明されるべきです。

表3：ステークホルダーに開示する人工知能モデルの主要な側面

出典：CIOMS Working Group XIV

<b>想定用途</b>	各AIモデルの想定用途は明確に定義し、関係者に伝達する必要があります。これには、自由記述テキストにおける有害事象認識、シグナル検出、症例トリアージなど、モデルが支援または実行することを目的としたPV業務を具体的に示すことが含まれます。
<b>人間-コンピュータ相互作用</b>	人間とAIモデルとの相互作用の水準および形態を明確にする必要があります。これには、AIモデルが自律的に実行されるのか、ヒューマン・イン・ザ・ループまたはヒューマン・オン・ザ・ループが存在するのか（およびそれらに求められる能力要件）、あるいは下流工程の専門家に対する意思決定支援を目的とするのかを示すことが含まれます。
<b>モデルアーキテクチャ</b>	AIモデルの種類および基本的な構造を開示する必要があります。例えば、ルールベースモデル、線形モデル、特定のニューラルネットワーク構造、あるいは複数の機械学習モデルを組み合わせたアンサンブルやマルチエージェントシステムなどが該当します。また、ニューラルネットワークの種類や深さなど、モデル構造に関する関連情報も共有する必要があります。
<b>モデルパラメータ</b>	少なくとも、AIモデルの判断に影響を与える主要な予測因子または特徴量が既知である場合には開示する必要があります。可能であれば、外部による再現および性能評価を可能にするため、モデルの重みやパラメーター式を共有することも考えられます。GenAIモデルを基盤とするAIソリューションでは、事前定義されたプロンプトに加え、前処理および後処理の手順も明示する必要があります。
<b>説明可能性</b>	可能な場合には、AIモデルがどのような一般原理や論理に基づいて結果や予測に到達するのかを説明する必要があります。説明可能性が十分でない場合には、その事実を明示し、影響について議論する必要があります（説明可能性の節も参照）。
<b>学習データセット</b>	独自の機械学習コンポーネントの開発に使用された学習データセットに関する情報を開示する必要があります。これには、データ規模、対象範囲、アノテーション指針、品質保証方法、作成時期、および想定運用領域との整合性評価と使用正当化が含まれます。
<b>標準AIコンポーネント</b>	AIモデルに、事前学習済みMLモデル、ライブラリ、フレームワーク、またはデータセットなどの公開された標準コンポーネントが組み込まれている場合は、その旨を開示する必要があります。これには、使用した具体的なバージョン、アクセス日、およびカスタムのパラメータ設定を含める必要があります。
<b>許容入力</b>	AIモデルが想定する入力データの種類を明示する必要があります。これにより、出力の根拠を理解できるようになり、設計対象外のデータ入力を防止して出力の正確性および信頼性を維持できます。
<b>出力の種類</b>	AIモデルが生成する出力の種類を説明する必要があります。例として、リスクスコア、分類結果、アラート、自由記述テキスト、ならびに出力に対するAIモデルの確信度を示す指標などが挙げられます。

<b>既知の限界</b>	AIモデルの特性に起因する既知の限界について伝達する必要があります。例えば、考慮できない特徴や相互作用、既知のバイアス、過少代表集団などが含まれます（第8章「公正性と公平性」参照）。
--------------	---

他の開発者や研究者がAIモデルを完全に再現し、さらに改変して利用できるようにするため、組織によってはモデルのパラメータおよび重みの完全なセットを公開したり、ソースコードを共有したりすることを選択する場合があります。このような高度な公開性は、外部専門家による査読および検証を可能にし、モデルの信頼性に対する信頼を高めるとともに、イノベーションの促進にも寄与します。しかしながら、知的財産、競争上の優位性、あるいは大規模モデルの極めて高い複雑性に関する理由から、常に実現可能であるとは限りません。さらに、多くのステークホルダーにとって、複雑なAIモデルの生コードやパラメータへのアクセスだけでは理解の向上につながらない場合があり、前述した他のモデル透明性確保手段によって補完される必要があります。モデルの機能および有効性について有意義な理解を得るためには、実装において採用された根拠、前提条件、および主観的判断を理解することの方が重要となる場合があります。完全な科学的再現性を確保するためには、開発者は関連する参照データセット、少なくとも性能評価に使用されたデータセットを共有する必要がある場合もあります。しかしながら、ユースケースや関与するステークホルダーによっては、これはデータプライバシー原則と抵触する可能性があります。

## 6.4. 説明可能性

透明性の一形態として、AIソリューションがどのような一般原理および論理に基づいて動作し、特定の出力に至ったのかを開示することが挙げられます。これは信頼の醸成に寄与し、影響を受ける個人が結果を理解し結果に影響を及ぼすことを可能にするとともに、その後の人間による意思決定を支援し、人間による監督および規制遵守を促進します。この文脈において、説明可能性（explainability）および解釈可能性（interpretability）は重要な概念であり、両者は部分的に重複しています。

ソフトウェアおよびシステムエンジニアリングにおけるソフトウェア試験のISO規格に含まれるAIベースシステム試験ガイドラインでは、説明可能性を「AIベースシステムが特定の結果にどのように到達したかを理解できる程度」と定義し、解釈可能性を「基盤となる（AI）技術がどのように機能しているかを理解できる程度」と定義しています。<sup>1</sup>

同様に、米国国立標準技術研究所（NIST）のAIリスクマネジメントフレームワークでは「説明可能性はAIシステムの動作を支える仕組みの表現を指し、解釈可能性はAIシステムの出力が設計された機能目的の文脈においてどのような意味を持つかを指す」と述べられています。<sup>2</sup>

OECDの透明性および説明可能性原則1.3では、次のように示されています。<sup>3</sup>

「説明可能性とは、AIシステムの結果によって影響を受ける人々が、その結果がどのように導かれたかを理解できるようにすることを意味する。これは、AIシステムの結果によって不利益を受けた人々が、その結果に異議を唱えることを可能にするため、理解しやすい情報を提供することを含み、とりわけ実行可能な範囲において、結果に至った要因および論理を提示することを意味する。」

本報告書の文脈では、これと同様の考え方を採用し、説明可能性という用語を、特定の結果に至った要因および論理、またはAIソリューションの一般的な動作に関与する要素を人間がどの程度理解できるかを示す、より広い概念として使用します。

異なるPVユースケースにおける説明可能性の役割を示す具体例は、付録3に示されています。

### 6.4.1. 説明可能性の利点

説明可能性には、以下のような利点があります。

- AIシステムの出力を理解し文脈の中で解釈できるようにすることで、ステークホルダーの信頼醸成に寄与します。
- AIシステムの出力によって影響を受ける個人が、その結果に異議を唱えたり影響を及ぼしたりすることを可能にします。
- AIシステムの出力を基盤として、またはそれを統合して行われる人間の意思決定を支援し、迅速化します。<sup>4,5</sup>
- エンドユーザーが検討すべき科学的仮説の提示につながります。すなわち、説明に含まれる薬剤、疾患、人口統計学的要因などの個別または組み合わせの特徴は、薬物有害反応のシグナルや薬剤—疾患相互作用として評価に値する可能性、さらには薬物有害反応の生物学的機序の示唆を提供する場合があります。<sup>6</sup>
- AIシステムに関するより完全な文書化、監査、および人間による監督を可能にします。
- 人間の意思決定、AI出力、およびその判断根拠となった説明を保持し検証できる場合には、特に規制遵守への貢献が期待されます。
- 潜在的なバイアスや偶発的の相関の可能性などの問題を明らかにすることで、トラブルシューティングを容易にします。<sup>7,8</sup>
- 同一データで学習した異なるモデル間で性能差が生じる要因を明らかにすることにより、モデル評価およびモデル選択に寄与します。

前述の定義に照らすと、PVにおけるAIシステムの出力に異議を唱える可能性があり説明可能性を必要とする主体は、一般市民よりも、PVプロセスに直接関与するステークホルダーである可能性が高いと考えられます。<sup>9</sup> これには、AIと直接相互作用するPVおよびQA担当者、AIシステムを構築または維持する開発者、さらには監査を行う規制当局が含まれます。PVプロセスに関与するさまざまなステークホルダーが説明可能性から得られる具体的な利点の例は、付録4に示されています。

### 6.4.2. 内在的説明可能性と事後的説明可能性

複雑性が限定されたAIモデルは、本質的に説明可能である場合があり、その出力の根拠をモデル構造やパラメータを直接確認することで導き出すことができます。<sup>9</sup> これはアンテホック (ante-hoc) 説明可能性とも呼ばれます。例としては、低次元の決定木 (decision tree)、ルールベース分類器、回帰モデルなどが挙げられます。

これに対し、複雑な構造を持つディープニューラルネットワークなど、人間が全体を把握または理解することが困難な多数のパラメータを有する不透明なAIソリューションに対しては、事後的説明可能性 (post-hoc explainability) を得ることを目的とした研究分野が発展しています。このような手法では、AIソリューションの上位に別の方法論や技術層を適用し、<sup>10</sup> 既に生成された特定の出力の根拠を追跡し説明します。いくつかの事後的説明可能性手法では、特徴量の相対的重要度を推定することで複雑なAIモデルの出力を説明し、別の手法では、特定の出力を変化させるために必要な特徴量の最小変更量を特定します。また、特定の出力の局所的な文脈に対して、より単純で本質的に解釈可能なモデルを当てはめることで説明を行う方法も存在します。本報告書作成時点で利用され

ている具体的手法の例については、[付録4](#)を参照してください。事後的説明可能性手法を用いる場合には、その手法自体もコンピュータ化システムに適用される規制要件に適合していなければなりません。すなわち、事後的説明可能性手法は目的適合性の観点から検証される必要があり、そのような手法を組み込んだプロセス自体もバリデーションされる必要があります。事後的説明は、データと予測結果との関係について近似的な理解を提供するものです。<sup>11</sup> その説明は不完全または部分的である可能性があり、必ずしも完全な説明を提供するとは限りません。

生成系大規模言語モデル（LLM）は、その出力に対するもっともらしい理由付けを生成するようプロンプトで指示することが可能です。しかし、このような理由付けは言語的生成物に過ぎず、モデル内部の計算過程へ直接アクセスできるものではなく、単なる事後的合理化である場合があります。PVでの利用においては、自由記述形式の「思考過程（chain-of-thought）」よりも、入力テキストの該当部分や根拠文書への明示的リンクに基づくエビデンス連結型の根拠提示が望まれる場合があります。また、LLMが生成した理由付けを監査証跡の一部として受け入れる前に、その忠実性（faithfulness）を確認する試験を要求することが望まれます。

### 6.4.3. 説明可能性に関連する課題

ステークホルダーは、想定されるユースケースにおいてどの種類の説明可能性が必要なのか、それが誰のためであり、どのような目的で必要なのかを批判的に検討し、さらに検討対象のAIシステムがそれを提供可能であるかを慎重に評価することが推奨されます。説明可能性はすべてのAIソリューションにおいて適切な目標であるとは限らない点にも留意が必要です。<sup>12</sup>

AIソリューションの出力における説明可能性の程度は、モデル選択を決定する唯一の要因であってはなりません。例えば、機械翻訳のような応用では、深層ニューラルネットワークが提供する高度な能力および性能向上に依存しており、本質的な説明可能性を必ずしも必要としません。説明可能性が限定的であることによる負の影響は、AIシステムの他の側面および性能に関する高い透明性を確保し、妥当性と頑健性、ヒューマン・オーバーサイトを強化することで軽減できます。<sup>13</sup> 同時に、説明可能性が必ずしも性能低下につながると想定すべきではなく、両者の間に必然的なトレードオフが存在すると考える必要もありません。<sup>14</sup>

また、AIモデルの出力に対する説明可能性は、妥当性・頑健性や公正性・衡平性に関する問題の特定に役立つ場合がありますが、説明可能性のみでシステムが目的適合であることを証明するものではなく、システムの信頼性を保証するものでもありません。<sup>15</sup> 説明可能性は特定の出力に至った要因を明らかにしようとするものですが、AIモデル全体の性能や公正性・衡平性を示すものではありません。例えば、本質的に説明可能なAIモデルで年齢が明示的特徴量として含まれていない場合でも、他の特徴量を介して特定の年齢層に対するバイアスが生じる可能性があります。実際には、説明が提示されることでステークホルダーがモデル出力を過度に信頼してしまう、いわゆる自動化バイアスを助長する可能性もあります。<sup>15</sup> また、説明可能性は透明性と同義ではありません。例えば、決定木のように本質的に説明可能なモデルであっても、組織が主要特徴や内部ロジックの開示を選択しない場合があります。

さらに、説明可能性の意味はすべてのステークホルダーにとって同一ではありません。モデル開発者にとって理解可能な説明が、他のステークホルダーには理解困難である場合があります。<sup>9</sup> そのため、説明はさまざまナリテラシー水準や教育背景を持つ人々にも理解可能である必要があります。<sup>10</sup> 説明を受け取る人間は、その内容を解釈し文脈化する必要があるため、自身のバイアスや認知的盲点が説明の活用に影響を与える可能性

についても認識しておく必要があります。関連して、最悪の場合には、不正確なAI出力に対してもっともらしい説明が付与されることで、一部の利用者が十分な批判的検討を行わずにその結果を受け入れてしまう可能性がある点にも注意が必要です。

## 6.5. パフォーマンスに関する透明性

特定のPV業務におけるAIモデルの評価済み性能に関する透明性は、AIモデルが実際の運用環境においてどの程度機能するかを示すものであり、モデル透明性によって提供される設計、実装、および意思決定プロセスに関する理解を補完します。これは理論的能力と実務上の有用性との間をつなぐ役割を果たします。AIモデルが現実的条件下でどのように挙動するかが明確でなければ、ステークホルダーはその適用適格性を十分に評価することができず、またその頑健性および妥当性に対して確信を持つこともできません。性能に関する透明性は、エンドユーザーから規制当局に至るすべてのステークホルダーが、AIモデルの強み、限界、および実際に導入される状況において期待される挙動を明確に理解できることを保証します。これは、患者安全および公衆衛生の保護を目的としてAIシステムが情報処理および意思決定を支援するPV分野において特に重要です。

関連するステークホルダーと詳細な性能評価結果を記録し共有可能とすることで、組織はAIシステムの強みおよび限界について明確性を提供します。これには定量的指標、定性的事例、およびベンチマークとの比較が含まれます。これにより、AIシステムに対する信頼および適切な依存を構築するために必要な文脈が提供されます。この透明性は、PV専門家および意思決定者による情報に基づく判断を可能にし、AIシステムが意図された適用範囲内で使用されることを保証するとともに、適応や特別な対策が必要となる領域の特定を支援します。さらに、モデルの追加改良や再学習が必要となる可能性のある領域を明らかにすることで、継続的改善も支援します。

これを支えるため、性能評価に使用されたデータに関する明確な文書化が存在する必要があります。これには、データ選定の根拠、取得方法、クリーニングおよび変換の方法ならびに欠損データや誤データを管理するためのプロセスが含まれます。

表4では、AIモデルの推定性能に関する透明性を確保するために開示すべき関連事項を示します。詳細については第5章「妥当性と頑健性」参照。

**表4：人工知能モデルの推定パフォーマンスに関する透明性を確保するために開示する関連する側面**

出典：CIOMS Working Group XIV

<b>評価範囲</b>	性能評価に使用した参照データセットの性質を記述します。想定される運用領域との既知の乖離（例：特定の医薬品、有害事象、患者集団の過剰または過少代表）についても明示します。関連情報には、使用したデータの種類およびその取得元を含めます。
<b>サンプリング</b>	参照データセットにおける陽性および陰性の出現頻度を記述し、それが想定使用条件とどのように関連するかを説明します。両者が異なる場合には、その差異を考慮して性能評価をどのように調整したかを説明します。性能評価においてデータ拡張を使用した場合には、その内容も記述します。
<b>参照基準</b>	性能評価において使用した分類カテゴリー（例：二値分類における陽性および陰性）の定義を開示します。参照基準作成時の人間による注釈の品質および一貫性を向上させるために用いたアノテーションガイドラインがあれば共有します。

人間の関与	テストセット作成に関与した評価者の資格を記述し、この段階で並行アノテーションや一致度評価を実施した場合はその内容を説明します。AIソリューションの運用にヒューマン・イン・ザ・ループが含まれる場合には、性能評価時に関与した担当者の資格も明示します。
要約指標	適切な場合には標準的な性能評価指標を提示し、独自指標を使用する場合にはその理由を説明します。提示にあたっては、想定される使用および運用領域に関連する意思決定閾値（例：偽陽性と偽陰性の現実的なバランス）に重点を置きます。複合指標（例：Fスコア）は、その構成要素（例：適合率（precision）および再現率（recall））と併せて提示します。
ベンチマーク	利用可能な場合には、関連するベンチマーク手法（人間の性能を含む）または標準ベンチマーク参照データセットとの比較結果を提示します。
サブセット解析および感度解析	性能評価において実施したサブセット解析または感度解析の結果を提示するか、実施していない場合にはその旨を明示します。
定性的レビュー	正しく分類された代表例および誤分類（偽陽性および偽陰性）の代表例を提示します。

## 第6章 – 参考文献

- International Organization for Standardization (ISO). ISO/IEC TR 29119-11:2020. Software and systems engineering — Software testing. Part 11: Guidelines on the testing of AI-based systems. Geneva: International Organization for Standardization; 2020. ([Abstract accessed 21 March 2025](#))
- National Institute of Standards and Technology. AI RMF Trustworthy & Responsible AI Resource Center. Gaithersburg, MD: National Institute of Standards and Technology. ([Webpage accessed 21 March 2025](#))
- Organisation for Economic Co-operation and Development (OECD). Transparency and explainability (Principle 1.3). [Internet]. Paris: OECD AI Policy Observatory; 2024. ([Webpage accessed 21 March 2025](#))
- Spiker J, Kreimeyer K, Dang O, Boxwell D, Chan V, Cheng C, Gish P, Lardieri A, Wu E, De S, Naidoo J. Information visualization platform for postmarket surveillance decision support. *Drug Saf*. 2020;Sep;43:905-915. <https://doi.org/10.1007/s40264-020-00945-0> ([Journal abstract](#))
- Botsis T, Dang O, Kreimeyer K, Spiker J, De S, Ball R. A Decision-Support Platform Powered by AI and Humans-in-the-Loop Boosts Efficiency and Assures Quality in FDA’s Pharmacovigilance. International Society of Pharmacovigilance, 23rd Annual Meeting 2024, Montreal, Canada. ([Webpage accessed 21 March 2025](#))
- Hauben M. Artificial intelligence in pharmacovigilance: Do we need explainability? *Pharmacoepidemiol Drug Saf*. 2022;Dec31(12):1311-1316. <https://doi.org/10.1002/pds.5501>. ([Journal full text](#))
- Ribeiro MT, Singh S, Guestrin C. “Why should I trust you?” Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York (NY): Association for Computing Machinery; 2016;Aug13:1135-1144. <https://doi.org/10.1145/2939672.2939778> ([Journal full text](#))
- Albahri AS, Duhaim AM, Fadhel MA, Alnoor A, Baqer NS, Alzubaidi L, Albahri OS, Alamoody AH, Bai J, Salhi A, Santamaria J. A systematic review of trustworthy and explainable artificial intelligence in healthcare: Assessment of quality, bias risk, and data fusion. *Inf Fusion*. 2023;Aug1;96:156-191. ([Journal full text](#)) <https://doi.org/10.1016/j.inffus.2023.03.008>
- Ghassemi M, Oakden-Rayner L, Beam AL. The false hope of current approaches to explainable artificial intelligence in health care. *Lancet Digit Health*. 2021;Nov1;3(11):e745-50. [https://doi.org/10.1016/S2589-7500\(21\)00208-9](https://doi.org/10.1016/S2589-7500(21)00208-9) ([Journal full text](#))

- 10 Reddy S. Explainability and artificial intelligence in medicine. *Lancet Digit Health*. 2022;Apr;4(4):e214–215. Cited by: Cutillo CM, Sharma KR, Foschini L, Kundu S, Mackintosh M, Mandl KD. Machine intelligence in healthcare - perspectives on trustworthiness, explainability, usability, and transparency. *npj Digit Med*. 2020;Mar;26;3(1):1-5. <https://doi.org/10.1038/s41746-020-0254-2> (Journal full text)
- 11 Pinheiro LC, Kurz X. Artificial intelligence in pharmacovigilance: a regulatory perspective on explainability. *Pharmacoepidemiol Drug Saf*. 2022;Dec1;31(12):1308-1310. <https://doi:10.1002/pds.5524> (Journal full text)
- 12 Royal Society. Explainable AI. [Internet]. London: Royal Society; 2019;Nov. (Website accessed 11 August 2024)
- 13 Ghassemi M, Oakden-Rayner L, Beam AL. The false hope of current approaches to explainable artificial intelligence in health care. *Lancet Digit Health*. 2021;Nov1;3(11):e745-50. [https://doi.org/10.1016/S2589-7500\(21\)00208-9](https://doi.org/10.1016/S2589-7500(21)00208-9) (Journal full text)
- 14 Rudin C, Radin J. Why are we using black box models in AI when we don't need to? A lesson from an explainable AI competition. *Harv Data Sci Rev*. 2019;Nov22;1(2):1-9. <https://doi.org/10.1162/99608f92.5a8a3a3d> (Journal full text)
- 15 Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. *J Am Med Inform Assoc*. 2012;Jan1;19(1):121-127. <https://doi.org/10.1136/amiajnl-2011-000089> (Journal full text)

# 第7章

## データプライバシー

### 原則

データプライバシーとは、個人が自身の個人情報の収集、保存、共有、および利用の方法を管理する基本的権利を指します。これは、生物医学研究の実施における基盤的原則である「人格の尊重 (respect for persons)」の一側面です。法令、規制、およびガイダンス文書は、医薬品安全性監視 (PV) の文脈において、機微性が高く個人識別につながり得る健康データの機密性、匿名性、自律性、および管理を維持することを目的とした各種措置を定めています。

### 主要メッセージ

- 保護対象データを含み得るPVにおけるAIの適用では、人間を対象とする研究活動に適用される標準的原則を考慮する必要があります。
- PVにおけるAIアプリケーションの利用では、データプライバシー要件に対応する適切な保護措置が確実に講じられるよう、追加的な配慮が求められます。
- 日常的なPV活動におけるAI利用に最も関連する倫理原則は、データプライバシー、公正性 (fairness)、および衡平性 (equity) です。
- PV専門家は、GenAIがデータプライバシーを損なうリスクを高め得ること、また機械学習 (ML) がバイアスを増幅し得ることにより、既存の規制遵守手順の再評価が必要となる可能性を認識する必要があります。

## 7.1. はじめに

データプライバシーは1世紀以上前から暗黙の法的権利として認識されてきましたが、<sup>1</sup>このテーマが国際的に正式な注目を集め始めたのは1970年代になってからです。コンピュータ技術の進歩により、それまで紙媒体に依存していた大量のデータの収集、整理、および評価が大規模に可能となりました。公的機関による個人データの収集、保存、共有の方法を規制する法律が存在しなかった状況の中で、1970年に最初のデータプライバシー法が制定されました。<sup>2</sup>米国では、収集されたデータの不適切利用に対する社会的懸念を背景として、1974年に米国プライバシー法 (US Privacy Act) が制定され、<sup>3</sup>連邦機関の記録システムに保持される個人情報の収集、維持、利用、および開示を規律する公正情報実務規範 (Code of fair information practice) が確立されました。同様の問題は、国境を越えて大量の個人データが移転されることへの懸念も引き起こし、国際貿易においてデータプライバシーを保護するための最初の国際的ガイドラインの策定につながりました。<sup>4</sup>本CIOMS作業部会報告書と同様に、OECDガイドラインも一連の中核的原則を提示していますが、その目的は政府、企業、および消費者代表を支援し、商取引を促進するためのデータ移転を可能にしつつ、個人データのプライバシー保護を確保することにあります。その後数十年にわたり、これらのガイドラインは多くの後続のデータ保護規制・法律に影響を与えました。例えば、1996年の医療保険の携行性と責任に関する法律 (Health Insurance Portability and Accountability Act: HIPAA、Pub. L. 104-109) や2016年一般データ保護規則 (General Data Protection Regulation: GDPR) が挙げられますが、これらについては本章後半で説明します。付録2に示されているとおり、近年の主要な国内外報告書においても、AIの一般利用および医薬品開発における利用の双方に関して、データプライバシー保護の考慮事項が明確に位置付けられています。

## 7.2. 倫理的配慮

AIの利用におけるデータプライバシーへの懸念は広く認識されていますが、本報告書発行時点では、PVにおけるAI利用に特化した倫理的考慮事項については十分な検討が行われているとは言えません。<sup>5</sup> WHOガイダンス「Ethics and Governance of Artificial Intelligence for Health」など、倫理とAIに言及する多くの文献では、<sup>6</sup> 1979年のベルモント・レポート（Belmont Report）で初めて体系的に示された基本原則が強調されています。<sup>7</sup> 同レポートは臨床研究および行動科学研究の倫理的枠組みを提供する目的で作成されましたが、その後、PV活動の一部にも適用されるようになりました（ただし、PVの大部分は通常「研究」とは見なされない点に留意が必要です）。

ベルモント・レポートでは、人間を対象とする介入研究および行動研究の基盤となる3つの基本原則が示されています。それは、「人格の尊重（respect for persons）」「善行（beneficence）」「正義（justice）」です。

人格の尊重とは、個人が研究への参加を自らの意思で決定できるという原則を指します。臨床研究においては、インフォームド・コンセントがこの原則の代表的な適用例です。PVでは、この原則はデータプライバシーおよび自己情報をコントロールする権利として適用され得ます（ただし、PV目的では法令により一定の制約が生じる場合があります）。善行（beneficence）とは、研究が参加者への利益を最大化し、危害を最小化するように設計・実施されるべきという原則を指します。PVに適用する場合、この原則は継続的なベネフィット・リスク評価として体现されます。正義の原則は、公正性および差別の排除に焦点を当てます。PVでは、公正性と衡平性（fairness and equity）として適用され、PVから得られる知見の利益が集団間で公平に分配されることを意味します。

ベルモント・レポートの本来の目的は、臨床研究および行動研究に適用される普遍的原則を確立することであり、公衆衛生活動を対象としたものではありませんでした。しかし、その後数十年の間に、公衆衛生分野は拡大し（例えば、1970年代には存在しなかったデータアクセスや研究手法の発展により）、疾病サーベイランスやPVなどの分野にもその原則が適用されるようになりました。例えば、製造販売後安全性試験（post-authorisation safety studies）は製品承認条件として求められることが多く、リアルワールドデータ（RWD）を用いてリアルワールドエビデンス（RWE）を生成する場合があります。CIOMS作業部会XIIIの報告書では、規制の意思決定におけるRWDおよびRWEが扱われています。<sup>8</sup>

PVにおけるAI応用（例：学習、妥当性評価、一般化可能性評価）に適用する場合、倫理的考慮事項として特に重要なのは、「人格の尊重<sup>6</sup>」に由来するデータプライバシーと「正義<sup>6</sup>」に由来する公正性と衡平性です。PVにおける公正性とは、差別的でない慣行を確保し、製品に曝露された集団を適切に代表する知見を得ることを意味します。また衡平性はPVの利益が広く共有されることを保証するために不可欠であり、このテーマは次章でさらに詳しく検討されます。

多くの国では、個人のデータプライバシー権を保護する法律が制定されています。これらの法律は、個人データは保護されるべきであり、そのためには個人へのリスクを低減する仕組みと、データを利用する主体の説明責任を求める仕組みが必要であるという共通原則を共有しています。代表的な例として、米国で適用される1996年医療保険の携行性と責任に関する法律（HIPAA）および欧州連合（EU）で適用される2016年一般データ保護規則（GDPR）が挙げられます。本章では、これらを例として、データプライバシー規制の共通点と相違点、およびPVにおけるAI適用への影響について説明します。

## 例：医療保険の携行性と責任に関する法律（HIPAA）

名称が示すとおり、HIPAA（1996年）は当初、医療保険データ<sup>9</sup>を対象としており、医療情報がアナログからデジタルへ移行する過程においてデータプライバシーを確保することを目的として策定されました。HIPAAはまた、医療分野における効率性向上を目的として、医療データに関する管理上の標準も導入しました。その後、医療分野におけるデジタル技術（例：電子健康記録〔EHR〕）の急速な普及および研究その他の目的で電子データを活用しようとする動きの高まりを受け、患者のプライバシーおよび安全性を保護しつつ管理効率を確保する基準のもとでEHRの利用を支援する追加規制が整備されました（HIPAA プライバシールール〔2000年〕、セキュリティルール〔2003年〕、Health Technology for Economic and Clinical Health〔HiTech〕法の Breach Notification Rule〔2009年〕）。HIPAA プライバシールールは、「対象事業体（covered entities）」と呼ばれる適用対象組織による個人の保護対象健康情報（protected health information：PHI）の利用および開示を規定するとともに、個人が自身の健康情報の利用方法を理解し管理するためのプライバシー権に関する基準を定めています。このプライバシールールの主要な目的は高品質な医療の提供および公衆の健康と福祉の保護に必要な健康情報の流通を可能にしつつ、個人の健康情報が適切に保護されることを確保することです。

HIPAAは、健康データの機密性、完全性および可用性を重視し、規制対象組織に対して特定の目的を達成するために必要最小限の範囲にPHIの利用、開示および要求を限定するよう合理的努力を求めています。また、患者がPHIへアクセスし修正を求める権利も規定しています。患者の機密性を保護するため、HIPAAは個人識別に使用され得るデータの種類を定義し、18種類の固有PHI識別子を特定しています。この一覧は、医療そのものとは直接関係しない場合であっても個人識別につながり得る一般的データの広い範囲を示しており、氏名、州より細かい地理的区分、生年を除く日付情報（例：誕生日）、電話番号、FAX番号、電子メールアドレス、社会保障番号、診療記録番号、保険加入者番号、口座番号、証明書・免許番号、車両識別情報、機器識別情報、ウェブURL IPアドレス、生体識別情報（例：指紋）、顔写真、ならびに個人の特定につながり得るその他の固有識別子が含まれます。これらの識別子が記録から除去された場合、その記録はPHIを含まないものとみなされ、HIPAAによる制限を受けることなく利用または開示することが可能となります。

公衆衛生においては、しばしば社会全体の利益と個人の権利との均衡が求められます。米国において使用が承認された医薬品の安全性、有効性および品質を確保するという社会的要請を踏まえ、申請者が実施する通常の医薬品安全性監視（PV）活動は、PHIの開示および利用に関する患者の許可を求めるHIPAAの一部要件から一般的に免除されています。医薬品は米国FDA規制のもとで管理されており、これには、とりわけFDA規制対象製品の品質および安全性の監視が含まれ、その一環として有害事象（AE）報告、製品追跡、回収、市販後監視が実施されます。もっとも、特定のPV活動がHIPAA要件の一部から免除される場合であっても、データプライバシー保護は引き続き適用されます。これには、最小必要データ原則（PV責務を果たすために必要不可欠なデータのみを収集すること）、可能な場合のデータの匿名化または識別子除去、PHIへの不正アクセス・利用・開示を防止するための技術的・管理的・物理的保護措置の実施、ならびに対象事業体がベンダーや提携先を利用する場合のBusiness Associate Agreementの締結要件が含まれます。米国内では、公衆衛生に関する責務は米国FDAおよび疾病管理予防センター（CDC, Atlanta）が公式権限のもとで担っています。米国FDAは、「ヒト用および動物用医薬品、生物学的製剤、医療機器等の安全性、有効性および安全保証を確保することにより公衆衛生を保護する」責務を負っています。一方、CDCは、「国外および米国内の健康、安全およびセキュリティ上の脅威から米国を守る」ことを使命とし、慢性疾患・急性疾患を問わず、公衆衛生上の脅威への対応を行います。これらは、「重要な科学的活動の実施および健康情報の提供」を通じて達成されます。

HIPAA プライバシールールは、疾病、傷害または障害の予防・管理を目的として法的権限を有する公衆衛生当局に対して、対象事業体が本人の許可なしにPHIを開示することを認めています。その場合でも、公衆衛生活動の目的達成に必要な最小限の範囲に合理的に限定することが求められます。<sup>10</sup>

これに対し、公衆衛生当局や民間部門とは異なり、学術機関が関与するPV活動は一般に研究活動として実施され、例えば製造販売後安全性試験（PASS）が該当します。この場合は異なる監督体制が適用され、研究が適切な倫理基準を満たしていることを保証するために機関審査委員会（IRBまたは倫理審査委員会）の審査が行われます。また、他機関との共同研究においては、データ利用契約の締結など、データプライバシー上の懸念を軽減する措置が求められます。これには、識別子除去データや限定データセットの利用、HIPAAおよびCommon Rule（45 CFR 46 Subpart A：人間を対象とする研究の倫理的実施に関するHHS方針）への遵守、ならびにその他適用される要件への適合が含まれます。

### 例：一般データ保護規則（GDPR）

GDPR（Regulation [EU] 2016/679）は、EUにおける個人データ保護を包括的に規定する制度であり、インターネット時代の黎明期にHIPAAと同時期に制定された旧データ保護指令（Directive 95/46/EC）を継承するものです。GDPRの適用範囲はHIPAAよりもはるかに広く、人と人とのあらゆる相互作用に関わる個人データの利用を対象とし、自動化された処理も含まれます。その根底には1950年の欧州人権条約における「すべての人は、その私生活および家族生活、住居および通信の尊重を受ける権利を有する」という原則があります。<sup>11</sup>

GDPRは、適法性、公正性、透明性、正確性および完全性、目的限定、データ最小化、機密性、保存期間制限といった原則を取り入れています。コンプライアンス確保はGDPRの重要な特徴であり、製薬企業を含む組織には遵守状況を監督するデータ保護責任者（Data Protection Officer）の設置が求められます。非遵守に対する制裁はHIPAAよりも大きく、最大で全世界売上高の4%に達する制裁金が科される場合があります。

データプライバシーを確保するための保護措置としては、暗号化（復号鍵なしではアクセスできないようにする）、可能な場合の匿名化、または仮名化（識別情報を仮名に置き換えて個人識別を困難にする）、さらにデータ処理に伴うリスクを特定・低減するためのデータ保護影響評価（Data Protection Impact Assessment）の実施などが挙げられます。HIPAAと異なり、GDPRには「忘れられる権利」が含まれており、個人は自身の個人データの削除を求めることができます。健康情報などの特別カテゴリーの個人データについては、GDPRの下では明示的の同意が必要となる場合があり、取得された同意は撤回可能です。

HIPAAと同様に、GDPRにおいても医薬品の安全性および有効性を確保する目的でAEの監視および報告を行うPV活動については、一定の保護措置が講じられることを条件として、同意なしで実施できる場合があります（ただし加盟国ごとの国内法の適用を受けます）。これらの責務により、通常GDPRで認められるデータ保護権（例：「忘れられる権利」）がPV活動において制限される場合があります。

その他の保護措置として、データ最小化原則の遵守や、個人データを保護するための管理的・技術的・組織的措置の実施が求められます。EUにおいて承認された医薬品の安全性、有効性および品質を確保する責務を果たすため、EMAはPV監督を実施する権限を有しており、その過程では「忘れられる権利」など特定のデータ保護権がPV活動において制限され得ることを前提としています。EMAはデータ利用において、データ最小化、目的限定、適法性、公正性および透明性の原則を重視しています。

GDPRは国際的なデータ移転に関して特別規定を設けており、EU市民（居住地を問わない）に関するデータを欧州経済領域（EEA）外へ移転する場合には制限が課され、適切なデータ保護水準を確保するための保護措置が求められます。EUにおける学術機関によるPV研究のデータプライバシー要件は米国の場合と概ね同様であり、IRBまたは独立倫理委員会（IEC）の監督の下、データ最小化および目的限定の原則の遵守が重視されます。また、EEA外へのデータ移転を伴う国際共同研究では、通常GDPR遵守を担保する契約条項を含む保護措置が必要となります。

### 7.2.1. その他のデータプライバシー関連法規

現在、US FDAおよびEMAに関連するデータプライバシー規制が最も広く参照されていますが、国ごとの規制差異は増加しており、二次利用データを用いた多国間AIモデル開発において特有の課題を生じさせている点に留意する必要があります。ブラジル、中国、ドイツ、日本における規制の比較は、この状況を示す一例です。

表5：ブラジルにおけるデータの二次利用に関するデータプライバシー規制

出典：CIOMS Working Group XIV

項目	ブラジル
適用法令	一般データ保護法（Lei Geral de Proteção de Dados：LGPD）、法律第13,709号（2018年）
健康データの分類	健康データは要配慮個人データに分類される。小児および青少年については追加的保護措置が適用される。
同意要件（研究およびPV）	原則として同意が必要。ただし例外あり（例：法的／規制上の義務、公的政策の実施、健康保護、認可機関による研究など）。
データの二次利用（例：RWE）	法的根拠により正当化される場合に許可（例えば、規制上の義務、公共政策、健康保護、研究）；同意の対象外となるANVISAの規制活動
匿名化基準	同意依存を低減する目的で匿名化および仮名化が推奨される。透明性、目的限定、データ最小化の原則が重視される。
越境データ移転	処理主体の所在地に関わらず、ブラジル国内個人のデータ処理に適用される。LGPD要件および保護措置を満たす場合、国際移転が可能。

項目	ブラジル
医薬品安全性監視 (PV) における例外	有害事象報告および技術的苦情は、規制/法的義務および健康保護を根拠として、ANVISAにより同意なしで取り扱われる。
生物医学的利用に関する規制当局ガイダンス	LGPDは全国的に施行。ANVISA令第1,184/2023号により個人データ保護方針（データインベントリ、安全対策、影響評価、契約遵守、データ保護文化等）が規定されている。
監督機関	国家データ保護機関（ANPD）。保健分野での実装はANVISAが担当。
主要参照文献	LGPD（法律第13,709号, 2018年）; <sup>12</sup> ANPD規則; <sup>13</sup> ANVISA令第1,184/2023号 <sup>14</sup>

表6：中国におけるデータの二次利用に関するデータプライバシー規制

出典：CIOMS Working Group XIV

項目	中国
適用法令	中華人民共和国個人情報保護法（PIPL）、中華人民共和国データセキュリティ法、中華人民共和国サイバーセキュリティ法
健康データの分類	「機微個人情報」に分類される
同意要件（研究およびPV）	原則として明示的同意が必要。厳格に解釈される
データの二次利用（例：RWE）	新たな同意取得または適切な匿名化が行われた場合に許容される
越境データ移転	有害事象（AE）報告は許容されるが、識別可能情報は最小化する必要がある
医薬品安全性監視 (PV) における例外	AE報告は許可されるが、識別可能なデータは最小限にしなければならない
生物医学的利用に関する規制当局ガイダンス	PIPLおよび健康データガバナンス規則を作成中
監督機関	中国国家インターネット情報弁公室（CAC）（サイバーセキュリティおよびデータ保護を監督）、国家衛生健康委員会（医療・健康データ基準の策定および実施を担当）

項目	中国
主要参考文献	個人情報保護法（2021）； <sup>15</sup> CAC健康データ規則案； <sup>16</sup> 国务院健康データ措置（2022） <sup>17</sup> ；データセキュリティ法； <sup>18</sup> サイバーセキュリティ法； <sup>19</sup> ネットワークデータ安全管理規定； <sup>20</sup> 越境データ移転安全性評価措置； <sup>23</sup> 越境データ流通促進および規制規定； <sup>21</sup> 基本医療衛生および健康促進法； <sup>22</sup> 人類遺伝資源管理条例； <sup>23</sup> 個人情報越境移転標準契約措置 <sup>24</sup>

表7：ドイツにおけるデータの二次利用に関するデータプライバシー規制

出典：CIOMS Working Group XIV

項目	ドイツ
適用法令	一般データ保護規則（GDPR：EU共通）；連邦データ保護法（BDSG：ドイツ）
健康データの分類	「特別カテゴリーの個人データ」（GDPR第9条）
同意要件（研究およびPV）	通常は同意が必要。ただし公共の利益（例：PV、RWE）に基づく例外あり
データの二次利用（例：RWE）	法的根拠（公衆衛生、科学研究等）が存在する場合、適切な保護措置の下で許可
非識別化基準	仮名化（pseudonymisation）を推奨；広範な再利用には完全匿名化を要求
越境データ移転	十分性認定国への移転、または標準契約条項（SCC）／拘束的企業準則（BCR）により許可
医薬品安全性監視例外	公衆衛生または法的義務に基づき、同意要件から明示的に除外
生物医学利用に関する規制当局ガイダンス	ドイツ各州データ保護監督機関（DPAs）および欧州データ保護会議（EDPB）
監督機関	ドイツのDPAおよび欧州データ保護委員会（EDPB）
主要参考文献	GDPR（EU規則2016/679）；EDPB Guidelines 03/2020；EMA GVP Module VI；BDSG（ドイツ）

表8：日本におけるデータの二次利用に関するデータプライバシー規制

出典：CIOMS Working Group XIV

項目	日本
適用法令	個人情報保護法（APPI）
健康データの分類	「要配慮個人情報」
同意要件 （研究およびPV）	原則として同意が必要。ただし仮名加工情報は公共利益または研究目的で利用可能
データの二次利用 （例：RWE）	仮名化／匿名化および研究目的の明示により利用可能
非識別化基準	匿名加工情報および仮名加工情報を認識（後者は引き続き規制対象）
越境データ移転	十分な保護水準にある国（EU、英国等）への移転は可能。それ以外は同意または契約措置が必要
医薬品安全性監視 例外	規制上の義務に基づき、有害事象報告は同意なしで実施可能
生物医学利用に 関する規制当局 ガイダンス	APPI下における臨床研究およびPVに適用される厚生労働省指針
監督機関	個人情報保護委員会（PPC）
主要参考文献	個人情報保護法（2020年改正〔訳注：原文作成時の最新改正〕）； <sup>25</sup> 医療分野の研究開発に資するための匿名加工医療情報及び仮名加工医療情報に関する法律〔訳注：次世代医療基盤法〕； <sup>26</sup> 医薬品の製造販売後の調査及び試験の実施の基準に関する省令（GPSP）； <sup>27</sup> 人を対象とする生命科学・医学系研究に関する倫理指針 <sup>28</sup>

### 7.3. 実際の配慮 データプライバシーのサポート

これらの例が示すように、通常の医薬品安全性監視（PV）活動を実施するために必要な枠組みの中で、適切なデータプライバシーを支援する規制が存在します。HIPAAで列挙されている18種類の固有識別子の一覧は、個人を特定するために利用され得るデータの種類が非常に広範であることを示しています。HIPAAおよびGDPRの導入以降、データを匿名化するためには追加的な対策が必要となる場合があることが認識されるようになりました。

規制産業である製薬企業は、自社製品が承認されているすべての国におけるデータプライバシーおよび報告要件を遵守しなければなりません。例えば、EMAはGVPへの遵守に加えて、GDPRに基づくデータ保護原則への適合を求めています。コンプライアンスを確保するためには、国ごとに進化する規制への対応、企業を支援するベンダー（場合によっては個別企業のPV活動の一部を実施する）の監督、さらに複数企業による配合剤共同開発のようなビジネスパートナーシップへの配慮が必要となります。規制当局ごとに患者情報の報告要件が異なる場合があり、各国要件への適合を確実にするためには個別対応や追加的な監督が求められます。例えば、オーストラリアでは公正性・衡平性を支援する目的で民族情報の報告が求められる一方、フランスでは差別への懸念からこれが禁止されています。

世界的なデータプライバシー要件への適合を支援するため、第三者（ベンダーやパートナー）との契約には、プライバシーに特化した条項や文言が含まれます。米国では、対象事業体がベンダーまたはパートナーと締結する契約にはBusiness Associate Agreementが必要となります。GDPRの下では、多国籍企業が国境を越えて企業内部で個人データを移転できるようにするため、拘束的企業準則（BCR）が導入される場合があります。BCRは法的拘束力を有し、EU当局の承認を必要とします。EUではさらに、製薬企業など一定の事業者に対して社内データ保護責任者の設置が義務付けられており、追加的な監督層が設けられています。世界的には、データ侵害に対する結果として、データ保護当局への通知義務から重大な罰金や制裁に至るまで、さまざまな措置が存在します。

### 7.3.1 医薬品安全性監視において人工知能が用いられる場合にデータプライバシーを維持する上でのリスク

AIの利点の一つは、日常的なPVデータ（例：ICSR）を大量に、より効率的に処理できるようにすることです。さらに、オープン型であれクローズド型であれ、LLMは、本来であれば起こらなかった、あるいは実行が困難であったデータソース同士の計画的（または非計画的）な連結を、ほぼ瞬時に可能にします。以下で述べるとおり、これらのリスクは、クローズドLLMと比べてオープンLLMの方が実質的に大きくなります。GenAIモデルは、外挿、すなわち本来であれば認識されなかった可能性のあるパターンの検出にも有用です。これらの特性は、現在のデータプライバシー保護手段が、識別情報が除去されたデータの再識別を防ぐのに十分であるのかという問いを提起します。

#### 識別情報除去措置の十分性

1990年（HIPAAの6年前）、米国の研究者は国勢調査データを用い、容易に入手可能な3つのデータ要素、すなわち5桁の郵便番号、性別、生年月日にもとづいて、米国人口の87%を同定できることを示しました。これは、個人を一意に同定するのに必要なデータ点のごく少数で足りることを示しています。<sup>29</sup>その後、郵便番号はHIPAAにおいてPHIとして分類されましたが、一般にアクセス可能なデータ点がわずかでも、データプライバシーが損なわれ得るという点は変わりません。

カナダのオンタリオ州にある小児病院のデータを用いた研究では、識別情報が除去された薬局データにもとづく個人の再識別リスクは、入院日と退院日を入院した四半期と年に置き換えるなど、選択したデータ要素の値の精度を下げることにより、最小化でき、場合によっては排除できることが示されました。しかし、データ要素の値について許容される一般化の最大程度は、再識別リスクおよび患者プライバシー侵害リスクだけでなく、意図された解析も正式に検討した上で決定しなければなりません。なぜなら、適切な精度がなければ当該解析を実施できない場合があるからです。<sup>30</sup>

EMAおよびHealth Canadaは現在、医薬品承認プロセスの一環として臨床試験報告書の公開共有を求めています。データ匿名化の標準も発出されています。これらの標準を適用し、研究者らは、非ステロイド性抗炎症薬の臨床試験報告書に関連する再識別リスクを評価し、疑われる症例を、正確に一致する可能性にもとづいて等級付けしました。<sup>31</sup>著者らは、500症例をレビューしたうち一致が疑われる症例が6件あることを見出し、さらに一致を同定する作業は時間を要すること（1症例あたり24.2時間）を観察しました。再識別の手がかりとしてはソーシャルメディアおよび死亡記録が最も有用でしたが、再識別が成功したかどうかは不確かでした。EMAが設定し<sup>32</sup>、Health Canadaが受け入れている再識別リスクの閾値（確率 $\leq 0.09$ ）にもとづき、著者らは、既存の匿名化ガイダンスは十分な水準のデータ保護を提供するのに足りる（ゼロではないが、信頼度の低い再識別

が少数にとどまることとして定義)と結論づけました。しかし同時に、この所見は希少疾患の試験には当てはまらない可能性があること、また匿名化において定量的手法ではなく定性的手法を用いる試験にも当てはまらない可能性があることも指摘しました。AIの急速な進歩により、その研究(2020年公表)で報告された再識別の再現に要する時間は、本報告書の時点までに短縮している可能性が高く、今後も短縮していくと考えられます。

## データ侵害のリスク

識別情報が除去されたデータが再識別されることによる潜在的な影響は、世界各地で発生している多数のデータ侵害事例によってさらに増幅されています。現在のAIは急速に進歩しており、AI手法がより高度化するにつれて再識別のリスクも増大すると考えられます。以下に、比較的最近発生したデータ侵害事例(当時は現在ほどAIが発達していなかった)と、それに伴う潜在的影響の広がりを示すいくつかの例を挙げます。

- 米国の金融企業に対する意図的攻撃により、1億件を超える顧客口座およびクレジットカード申請情報へのアクセスが発生しました。<sup>33</sup>
- 外国政府による政治的動機が疑われる国際的サイバー攻撃により、米国の信用情報機関が侵害され、米国民のほぼ半数に相当する氏名、生年月日、社会保障番号が流出しました。この攻撃は、AIを用いて米国政府関係者を標的とすることを目的としていたとされています。<sup>34</sup>
- 政治的対立相手を失墜させる目的で行われた国内の意図的データ侵害として、アジアの医療保険プランに対するサイバー攻撃が発生し、150万件の患者記録が流出しました。<sup>35</sup>
- インド政府が保有する生体認証情報およびその他の個人データが、12億人分の記録を含むデータベースから意図せず公開されました。<sup>36</sup> この事例では、犯罪グループが侵害を悪用し、個々の患者記録を販売しました。約10万人のデータが実際にアクセスされたことが確認されています。

これらの事例はいずれもGenAIが広く利用される以前に発生したもののですが、急速に進歩しているGenAIは、公開情報源と不正取得されたデータを効率的に結び付ける可能性を有しており、再識別リスクの増大およびデータプライバシー侵害の深刻化につながるおそれがあります。

## 個人による個人データ保護の責任

データプライバシー規制への適合を目的とした組織的な対策に加え、個人自身も自らのデータを保護する役割を担います。この責任は、スマートフォンやウェアラブル機器などのデジタルツール、各種アプリケーション、ならびにAI支援翻訳ツールやGenAIなどのソフトウェアの急速な普及に伴い、ますます重要になっています。これらは、十分に保護されない形で個人データが開示される機会を生み出す可能性があるためです。多くの場合、GDPRなどにより個人のデータプライバシーを保護する法的枠組みが存在しますが、それでもプライバシー侵害が生じる余地は残っており、とりわけGenAIの利用において懸念が高まっています。データ利用方針を通知する一般的な手段としては、利用規約(例:エンドユーザー使用許諾契約[EULA])、プライバシー通知、場合によっては個人データ利用に対する明示的同意の取得などがあります。利用者は、問い合わせ時に氏名や電話番号などの個人情報を共有してしまいがちですが、その結果を十分理解していない場合があります。特にオープン型GenAIツールを利用する場合、これらのデータはもはや私的情報として保持されない可能性があります。さらに、最近の疾病や交通事故など、状況依存的な情報を共有することで個人が特定され得る場合があります。<sup>37</sup>

また、GenAIが生成した内容をソーシャルメディア等で共有する際、利用者が意図せず個人識別情報を含めてしまう可能性もあります。こうした情報は、プラットフォームのデータプライバシー方針によっては、直接個人情報を入力していなくても個人特定に利用される可能性があります。GenAIが急速に進化していることを踏まえると、GenAIにおけるデータプライバシーを保護する規制も同様に進化していく必要があります。

## 医薬品安全性監視における大規模言語モデルの使用に伴うデータプライバシーへの潜在的リスクと、軽減のためのアプローチ

原則として、既存のデータプライバシー規制（またはGDPRのような法令）は、PVにおけるAIアプリケーションで使用されるデータ保護の基盤となるはずですが。モデル開発では、学習および実行のために、データサイエンティストや機械学習（ML）エンジニアがPVデータ（ICSRなど）を使用することが求められる場合があります。関与する当事者（製薬企業およびベンダーを含み得ます）は、保護対象データにアクセスする可能性があります。アクセス可能な集団が拡大することで漏えいリスクが生じます。関与する全ての当事者は、データプライバシー要件を認識しているべきです。大規模言語モデル（LLM）では、これらのモデルの基盤となる仕組みにより、本来は起こりにくいはずの再識別が一部生じ得るため、リスクが高まる可能性があります。クローズドLLMを維持する組織は、プロンプトならびにモデルに含まれるデータを管理できる場合があります。一方、オープンLLMモデルにはこのような保護策がなく、再識別リスクが増大します。これはこれらのモデルが、必ずしもデータプライバシー規制の適用範囲に入らない多様なデータソース（例えば、上記「個人による個人データ保護の責任」で述べた種類のデータを含む）にアクセスできるためです。漏えいは、データプライバシーへの配慮を回避するプロンプト、または個人データで学習されたモデルを通じて生じ得ます。さらに、上記のとおり、識別情報が除去されたと想定されるデータであっても（例えば、郵便番号、生年月日、性別、民族性が保持されている場合）、再識別が生じ得ます。PVでは、生年月日を含む基本的属性情報など、潜在的に識別可能で機微な情報のレビューが必要となります。これらは、医療・健康情報（医薬品およびワクチン曝露を含む）に加え、民族性、人種、性的指向、遺伝情報、生体情報、身体的特徴、生活習慣情報等の機微なデータ要素と関連し得るため、より強化された保護措置が求められます。

PVにおいて、GenAI（および場合によってはML）がもたらす課題の種類として、以下が挙げられます。

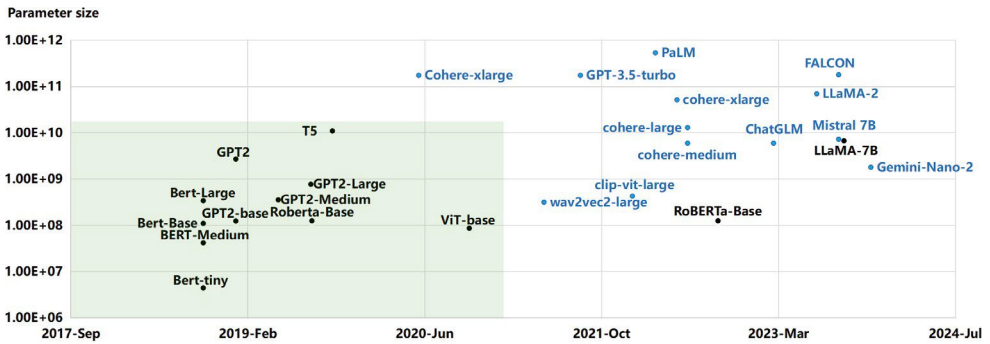
- 適用されるデータプライバシー要件に十分注意を払わないまま、オープンLLM内でアルゴリズムが開発される可能性があり、その結果、潜在的なプライバシーリスクが生じます。こうしたLLMがクローズドLLM内での使用に採用される場合、保護情報が開示される潜在的リスクがあります。
- クローズドLLM内では、無関係の目的でLLMに追加され得る様々な情報源にも注意を払うべきです。研究のために遺伝データが収集され（参加者の同意に基づく）、特定の解析のためにLLMへ追加される場合、元の同意の範囲外となる別目的で使用されないようにする措置が必要になります。一般に受け入れられている実務としては、当該データの追加利用について同意を取得する（データがLLMの一部となるため）か、または適切な管理（例：アクセス制御、入力および出力の監視によるデータ漏えいの軽減）を確保することが挙げられます。<sup>38,39</sup>
- GenAIプログラムは、国勢調査、人口動態統計、公衆衛生データなど、本来は分離しているデータソースを統合し得ます。これらが、識別情報が除去された健康記録データセット（例えば、市販後安全性調査のような継続中のPV活動の場面）と連結されることで、再識別の可能性が生じ得ます。

- その他のプライバシーリスクとして、以下が含まれ得ます。(i) データの永続性（すなわち、保持期間が経過してもAIシステム内にデータが保持され、削除されないこと）、(ii) データ主体の権利（例えば、AIシステムが学習に用いたデータセットを、システムがもはや「そのデータセット」として保持していない場合のアクセス権など）を遵守することの複雑さ、(iii) 透明性（例えば、モデルがデータ主体の認識なくデータスクレイピング（ウェブ等）により学習される場合）です。

GenAI/AIの各ユースケースは、地域の法令および関連する国際的に適用され得る法令を遵守しなければなりません。データプライバシー規制が緩い、または執行が不十分な環境では、リスクは増幅されます。小規模で早期に導入されたLLMは、より大規模で新しいLLMよりもデータプライバシーに関する精査を受ける傾向がありました（下図5参照）。理由としては、(1) より新しく大規模なLLMが一般に利用可能になっていないこと(2) プライバシー技術が、こうした新しく大規模なLLMに追従することが困難であったこと、が挙げられます。

図5：大規模言語モデルのプライバシー保護に関する研究の状態（2025年6月現在）

出典：Modified from Yan B et al 2025<sup>40</sup>



横軸は大規模言語モデル（LLM）の公開時期を示し、縦軸はパラメータ数の規模を示しています。青色のデータ点は、プライバシー保護に関する文献において十分な検討がなされていないLLMを示し、黒色のデータ点は、プライバシーに関する検討とともに研究されてきたモデルを示しています。緑色の背景は、プライバシー保護の向上に寄与し得るLLMの中心的なクラスターを強調しています。本論文公表時点では、近年導入されたより大規模なLLMは、初期の小規模モデルと比較して、データプライバシーの観点から科学文献における評価が十分に行われていない状況でした。265。

## 7.4. 結論

データプライバシーの権利は、ベルモント・レポートにおいて示された、人間を対象とする研究の保護に関する基本的倫理原則という確立された枠組みの中に位置付けられています。各国の規制当局は、データプライバシーを保護することを目的として、利用可能なデータの種類および必要な保護措置（データ最小化、匿名化、識別子除去、データ暗号化など）を示す要件を定めており、これらには不遵守時の罰則も含まれています。しかしながらデータプライバシー法の整備および個人情報を取り扱う組織における技術的対策の高度化にもかかわらず、データ侵害の試みおよび実際の侵害は、近年ますます

頻繁に、かつ極めて大規模（場合によっては1億人以上に影響）に発生しています。これは監督体制の不備に加え、サイバーセキュリティ対策を回避して安全なデータ源へ侵入する技術的進歩が存在することを示唆しています。

計算能力の継続的な向上、相互接続された大規模データベースの拡大、そしてGenAIの導入は、より多くかつ複雑な相互依存関係（例：事業提携、国際的ベンダーによるPV業務実施）を伴う、ますますグローバル化した医薬品安全性監視（PV）環境と並行して進行しています。この急速に進化するデータサイエンス環境において、PV専門家、規制当局、産業界、PV組織および学界にとって継続的な課題となるのは、個人データを適切に保護するためにデータプライバシー対策を監視し、定期的に更新し続けることです。

PVにGenAIを適用する際の潜在的リスクの一つとして、患者の再識別が挙げられます。これは、識別子除去データの具体性の在り方を再検討する必要性とともに、一部のデータ源が利用者の管理外に存在し得るオープンLLMに関連するリスクを示唆しています。

既存のデータプライバシーポリシーを整備し、関連規制を遵守することに加えて、AIをPVに適用する際のデータプライバシーリスクを軽減するための対応として、以下が挙げられます。

- 技術が急速に進展していることを認識し、例えば識別子除去措置が十分であることを確保するための継続的監視を行うこと。
- 再識別リスク低減を目的としてGenAI企業が導入するポリシーに注意を払うこと。<sup>41</sup>
- GenAIモデルに対してシステミックリスク軽減を目的とした要件を法規制が課す可能性があることを認識すること。<sup>42,43</sup>
- オンプレミスまたはプライベートクラウドでの運用、高度な匿名化技術、フェデレーテッドラーニング構成、差分プライバシー手法、ならびにAIサービス提供者との間でのデータ取扱い、保存および越境移転を規定する包括的な契約的保護措置を導入すること。
- オープンLLMとクローズドLLMがデータプライバシーに対して異なる課題をもたらすことを理解すること。機関内ファイアウォールで保護された環境においてクローズドLLMを運用し、第三者との共有に伴うリスクを慎重に評価することは、リスク低減に有効です。
- 報告書に含まれる個人情報が必要最小限であること、二次利用が当初の収集目的と整合していること、ならびにPVに関与するすべての主体（例：ベンダー）がデータ保護要件を遵守していることを確認するための監査を実施すること。PV業務が複数組織の協働ネットワークで実施される場合、最も監督体制の弱い組織がリスクとなり得ます。
- PV実務におけるLLM利用について、適切に訓練されたPV専門家による目的適切な使用が確保されるよう、アクセス管理および監督を行うこと。

## 第7章 – 参考文献

- 1 Warren SD, Brandeis LD. The right to privacy. *Harv Law Rev.* 1890;4(5):193-220. ([Webpage](#) accessed 27 April 2025)
- 2 Gesetz zum Schutz personenbezogener Daten (Hessisches Datenschutzgesetz – HDSG). Gesetz- und Verordnungsblatt für das Land Hessen I. 1970;S.61. ([Webpage](#) accessed 15 October 2025)
- 3 United States. Privacy Act of 1974, Pub. L. No. 93-579, § 552a, 88 Stat. 1896 (1974) (codified as amended at 5 U.S.C. § 552a). ([PDF](#) accessed 15 October 2025)
- 4 Organisation for Economic Co-operation and Development (OECD). Guidelines on the protection of privacy and transborder flows of personal data. OECD Doc. C(80)58/FINAL. Paris: Organisation for Economic Co-operation and Development; 1980;Sep 23. ([Webpage](#) accessed 15 October 2025)
- 5 Jain A, Salas M, Aimer O, Adenwala Z. Safeguarding patients in the AI era: ethics at the forefront of pharmacovigilance. *Drug Saf.* 2024;48:119-127. <https://doi.org/10.1007/s40264-024-01483-9> ([Journal abstract](#))
- 6 World Health Organization (WHO). Ethics and governance of artificial intelligence for health: WHO guidance. Geneva: World Health Organization; 2021. ([Webpage](#) accessed 25 March 2025)
- 7 U.S. Department of Health & Human Services (HHS). The Belmont Report: ethical principles and guidelines for the protection of human subjects of research. Washington (DC): U.S. Department of Health & Human Services; 1979. ([Webpage](#) accessed 21 March 2025)
- 8 Council for International Organizations of Medical Sciences (CIOMS). Real-world data and real-world evidence in regulatory decision making: CIOMS Working Group report. Geneva: Council for International Organizations of Medical Sciences; 2024. <https://doi.org/10.56759/kfxh6213> ([Full text](#))
- 9 United States, Department of Health & Human Services (HHS). Health information privacy: HIPAA for professionals. [Internet]. Washington (DC): U.S. Department of Health & Human Services; 2024. ([Webpage](#) accessed 21 March 2025)
- 10 United States, Department of Justice, National Security Division. Data Security Program [Final rule implementing Executive Order 14117]. 28 CFR Part 202. Washington (DC): U.S. Department of Justice; 2025;Apr 8. ([Webpage](#) accessed 22 September 2025)
- 11 GDPR.eu. What is GDPR, the EU's new data protection law? [Internet]. Brussels: GDPR.eu; 2024. ([Webpage](#) accessed 21 March 2025)
- 12 Brazil. Law No. 13,709 of 14 August 2018. Provides for the protection of personal data and amends Law No. 12,965/2014 (the Brazilian Internet Civil Framework) [Internet]. Brasília: Government of Brazil; 2018. ([Webpage](#) accessed 15 October 2025)
- 13 Brazil. Law No. 13,853 of 8 July 2019. Amends Law No. 13,709 of 14 August 2018 to provide for the creation of the National Data Protection Authority (ANPD) and makes other provisions [Internet]. Official Gazette of the Federal Government. Brasília: Government of Brazil; 2019;Jul 9. ([Webpage](#) accessed 15 October 2025)
- 14 Brazil. ANVISA Ordinance No. 1,184 of 17 October 2023. Establishes the Personal Data Protection Policy of the Brazilian Health Regulatory Agency (ANVISA) [Internet]. Brasília: Agência Nacional de Vigilância Sanitária; 2023;Oct17. ([Webpage](#) accessed 15 October 2025)
- 15 National People's Congress (China). Personal Information Protection Law of the People's Republic of China. Beijing: National People's Congress; 2021. ([Webpage](#) accessed 15 October 2025):
- 16 Cyberspace Administration of China (CAC). Official website of the Cyberspace Administration of China. [Internet]. Beijing: Cyberspace Administration of China; 2025. ([Webpage](#) accessed 15 October 2025)
- 17 National Health Commission of the People's Republic of China (NHC). Official website of the National Health Commission. [Internet]. Beijing: National Health Commission; 2025. ([Webpage](#) accessed 15 October 2025)
- 18 National People's Congress (China). Data Security Law of the People's Republic of China. Beijing: National People's Congress; 2021. ([Webpage](#) accessed 15 October 2025)
- 19 National People's Congress (China). Cybersecurity Law of the People's Republic of China. [Internet]. Beijing: National People's Congress; 2016;Nov7. ([Webpage](#) accessed 15 October 2025)
- 20 State Council of the People's Republic of China. Regulations on network data security management. [Internet]. Beijing: State Council of the People's Republic of China; 2024;Sep30. ([Webpage](#) accessed 15 October 2025)

- 21 State Council of the People's Republic of China. Regulation on the protection of non-public data. [Internet]. Beijing: State Council of the People's Republic of China; 2024;May30. ([Webpage](#) accessed 15 October 2025)
- 22 National People's Congress (China). Law of the People's Republic of China on Basic Medical and Health Care and the Promotion of Health. [Internet]. Beijing: National People's Congress; 2019;Dec28. ([Webpage](#) accessed 15 October 2025)
- 23 State Council of the People's Republic of China. Regulations on the Administration of Human Genetic Resources (2024 Revision). [Internet]. Beijing: State Council of the People's Republic of China; 2024. ([Webpage](#) accessed 15 October 2025)
- 24 National People's Congress (China). Standard Contract Measures for the Export of Personal Information. [Internet]. Beijing: National People's Congress; 2023;Feb22. ([Webpage](#) accessed 15 October 2025)
- 25 Japan. Act on the Protection of Personal Information [Internet]. Amended June 12 2020; Personal Information Protection Commission; 2020. ([Webpage](#) accessed 15 October 2025)
- 26 Japan. Act on Anonymized Medical Data That Are Meant to Contribute to Research and Development in the Medical Field [Internet]. Law No. 85 of 2017 (enacted May 9 2017; effective May 30 2018). Ministry of Health, Labour and Welfare & other competent ministries; 2017. ([Webpage](#) accessed 15 October 2025)
- 27 Japan. Ministerial Ordinance on Good Post-marketing Study Practice for Drugs (Ordinance No. 171 of 2004) [Internet]. Tokyo: Ministry of Health, Labour and Welfare; 2004. ([Full text](#) accessed 15 October 2025)
- 28 Ethical Guidelines for Medical and Biological Research Involving Human Subjects (Japan) Japan. Ethical Guidelines for Medical and Biological Research Involving Human Subjects [Internet]. Public Notice of MEXT / MHLW / METI No.1 of 2021. Tokyo: Ministry of Health, Labour and Welfare; 2021. ([Journal full text](#) accessed 15 October 2025)
- 29 Sweeney L. Simple demographics often identify people uniquely. Carnegie Mellon University, Data Privacy Working Paper No. 671. Pittsburgh (PA): Carnegie Mellon University; 2000;Jan;1-34. ([Full text](#) accessed 21 March 2025)
- 30 El Emam K, Dankar FK, Vaillancourt R, Roffey T, Lysyk M. Evaluating the risk of re-identification of patients from hospital prescription records. *Can J Hosp Pharm*. 2009;Jul;62(4):307-313. <https://doi:10.4212/cjhp.v62i4.812> ([Journal full text](#))
- 31 Branson J, Good N, Chen JW, Monge W, Probst C, El Emam K. Evaluating the re-identification risk of a clinical study report anonymized under EMA Policy 0070 and Health Canada Regulations. *Trials*. 2020;Dec;21:1-9. <https://doi.org/10.1186/s13063-020-4120-y> ([Journal full text](#))
- 32 European Medicines Agency (EMA), Guidotti T. External guidance on the implementation of the European Medicines Agency policy on the publication of clinical data for medicinal products for human use. Amsterdam: European Medicines Agency; 2018. ([Webpage](#) accessed 21 March 2025)
- 33 Starks T. U.S. charges Chinese military hackers with massive Equifax breach. [Internet]. Atlanta (GA): CNN; 2020. ([Article](#) accessed 21 March 2025)
- 34 Geller E. U.S. charges Chinese military hackers with massive Equifax breach. [Internet]. New York: The New York Times; 2020. ([Article](#) accessed 21 March 2025)
- 35 The Straits Times. Personal info of 1.5 million SingHealth patients, including PM Lee, stolen in Singapore's most serious cyber attack. [Internet]. Singapore: The Straits Times; 2020 ([Article](#) accessed 21 March 2025)
- 36 Jain M. The Aadhaar card: cybersecurity issues with India's biometric experiment. Seattle (WA): Henry M. Jackson School of International Studies, University of Washington; 2019. ([Article](#) accessed 21 March 2025)
- 37 Jamney V, Elkin PL. Re-identification risk in HIPAA de-identified datasets: the MVA attack. *AMIA Annu Symp Proc*.2018;Dec 5;2018:1329-1337. ([Journal full text](#) accessed 15 October 2025)
- 38 Ayyamperumal SG, Ge L. Current state of LLM risks and AI guardrails. [preprint]. arXiv. 2024;Jun16. <https://doi.org/10.48550/arXiv.2406.12934> ([Journal full text](#))
- 39 Asthana S, Zhang B, Mahindru R, DeLuca C, Gentile AL, Gopisetty S. Deploying privacy guardrails for LLMs: a comparative analysis of real-world applications. [preprint]. arXiv. 2025;Jan 21. <https://doi.org/10.48550/arXiv.2501.12456> ([Journal full text](#) accessed 22 September 2025)
- 40 Yan B, Li K, Xu M, Dong Y, Zhang Y, Ren Z, Cheng X. On protecting the data privacy of large language models (LLMs) and LLM agents: a literature review. *High-Confid Comput*. 2025;Feb 28;100300 <https://doi.org/10.1016/j.hcc.2025.100300>. ([Journal full text](#))
- 41 OpenAI. Usage policies. [Internet]. San Francisco (CA): OpenAI; 2025;Jan 29. ([Webpage](#) accessed 22 September 2025)
- 42 European Parliament, Council of the European Union. Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) (Article 51§ 1, a). *Off J Eur Union*. 2024;Jun 13;L 1689:1-145. ([Webpage](#) accessed 22 September 2025)
- 43 European Parliament, Council of the European Union. Annex XIII: criteria for the designation of general-purpose AI models with systemic risk (in Regulation (EU) 2024/1689, Artificial Intelligence Act). *Off J Eur Union*. 2024;Jun 13;L 1689:136-138. ([Webpage](#) accessed 22 September 2025)

## 第8章

# 公正性と衡平性

### 原則

公正性と衡平性（fairness and equity）〔訳注：3頁訳注参照〕は、公明正大、平等、非差別、多様性、正義、および法令遵守に対する認識とそれらの遵守を要求します。医薬品安全性監視（PV）における人工知能（AI）の便益は、関連するすべての集団およびグループに対して衡平に提供されるべきです。AIライフサイクル全体を通じて、不公平なバイアス、差別的慣行、不当な社会的福祉および環境への影響を回避し、軽減することが重要です。

### 主要メッセージ

- 公正性と衡平性に影響を与えるAIの開発および適用を考慮することが重要です。これらが欠如または不均衡である場合、AIソリューションによって十分に対象とされていないサブ集団に対する差別的な不利益、明示的バイアスによる負の影響、または不正確な結果の提供による性能低下が生じる可能性があります。
- バイアスが導入され得る領域については、可能な場合に軽減戦略を計画および実装し、潜在的な性能低下を減少させるとともに、十分に対象とされていない集団に対する差別的な不利益を回避します。
- 衡平性は、代表性のあるデータセットの評価などの措置を講じることにより促進され得ます。これにより、PVにおけるAIの適用結果（例：評価結果、安全性評価に用いられる集計データ出力）が、評価対象医薬品に曝露されることが想定される集団に対して適切なものとなるよう確保します。
- 適切な統計手法を用いて明示的または潜在的なバイアスを可能な限り検出・特定することは、リスク低減のための軽減措置の実施、AIの適用可能性および限界の把握、ならびに期待される性能受入基準の確立において重要です。
- 学習および性能評価に用いられる参照データセットについて、十分な代表性が確保されているかを精査し、可能な場合には関連するサブグループにおける性能評価を実施します。不十分な参照データは、公正性と衡平性の不足の主要な原因となることが多いです。
- 自発報告システムに固有の既知の制約により、個別症例安全性報告（ICSR）における公正性と衡平性は依然として限定的です。一部の国では他国よりも報告数が著しく多く、リアルワールドデータ（RWD）など分析に有用な追加情報も提供される傾向があります。その結果、十分に対象とされていない集団における実際の使用状況に関する理解はしばしば限定的となります。

## 8.1. はじめに

PVにおいては、AIの導入後も、プライバシー法やPV規制など既存の法令および規制の遵守は維持されなければなりません。変化している点は、AIの種類、データソース、および適用方法に応じて、公正性と衡平性にさまざまな程度で影響を及ぼし得る要因について、検討、ガバナンス、および軽減措置の必要性に対する認識が高まっていることです。

AIの利用に伴うすべての公正性と衡平性の概念、考慮事項、または負の影響がPVに固有であるわけではありません。公正性と衡平性に関する検討は困難であり、文化的差異、歴史的不平等、認識、社会経済的差異などの影響を受ける可能性があり、主観的に見える場合もあります。しかし、それはこれらの検討を行う必要性を否定するものではありません。バイアスを限定するためには、AIライフサイクル全体にわたり、すなわち構想および設計段階から実装、さらに本番運用中に至るまで、差別的リスクを低減するための意図的な対応が求められます。<sup>1</sup>

AIソリューションに内在するバイアスは一般的な問題であり、性能に影響を及ぼす可能性があります。すべての統計的バイアスが公正性と衡平性に対する負の影響を生じるわけではありません。PVにおいては、データ収集、選択、モデル開発、ならびにAIの設計・開発・利用に関与する人間の関与を通じて導入され得る不公平なバイアスに焦点を当てます。これらは不公平、差別、または不平等を引き起こす可能性があります。

本章では、AIの開発および利用が正義および法令遵守、個人の必須サービスへのアクセス、AIシステムおよび必要なエコシステムの導入・運用に必要な公的資源の不足、ならびに社会的福祉への影響に与える影響については扱いません。これらは重要な課題ではありますが、PVに固有のものではないためです。ただし、AIおよび必要なエコシステムへのアクセスは、PVに固有ではないものの、低・中所得国においては大きな障壁となり不平等や十分に対象とされない集団を生み出す可能性があります。また、PVにおけるAI導入に伴う労働力への影響については、本報告書では扱わず、第4章「ヒューマン・オーバースイト」4.3「従来の役割の変革」で議論されています。さらに、GenAIを含むAI利用の急速な拡大は、エネルギー需要の著しい増加および環境への影響を伴いますが、これらは社会全体に広く影響する課題であり、本報告書の対象範囲外とします。<sup>2</sup>

## 8.2. 公正性と衡平性の考慮事項と医薬品安全性監視

公正性と衡平性の原則は、AIシステムの利用により生じ得る差別的バイアスを特定し対処するうえで基本的なものです。PVにおいては、医薬品の安全かつ有効な使用を確保するため、有害事象を検出、評価、理解および予防するための積極的な対応が不可欠です。<sup>3</sup> AIソリューションをPVに適用する場合、高リスクのPV活動のために開発されるAIシステムによって生じ得る潜在的な危害を軽減するための予防的戦略を実装することが重要です。

AIシステムのライフサイクル全体を通じて、潜在的なリスク領域およびバイアスが導入され得る仕組みを特定し定量化するために、十分な評価および継続的なモニタリングが必要です。これは、PVにおけるAIシステム利用に伴う差別的バイアスを軽減するための戦略を定義する手段となります。バイアスの監視は、構想、開発、試験、およびソリューション導入後を含めて実施する必要があります。バイアス監視の頻度および適切な修正の必要性は、リスク評価、ソリューションの結果、ならびにモデルのバイアスおよび性能に影響を与え得る外部要因に基づいて定義されるべきです。

医薬品に関連する有害事象の検出、収集、評価、監視、理解および予防に関わる活動にAIを使用する際には、不公平な慣行または患者の不平等な取扱いにつながる可能性のあるバイアスが存在し得ることを認識することが極めて重要です。PV活動にAIを適用するPV組織は、当該ソリューションが定義された業務要件を満たし、患者安全活動を適切に支援し、人種、性別、年齢、社会経済的要因などに基づく排除によって、患者を不利な立場に置いたり、差別を生じさせたり、あるいは医薬品から得られる潜在的利益を享受できない状況を生じさせるなど、意図せず患者をリスクにさらすバイアスを導入しないことを確保する責任を負います。

## 8.3. 公正性と衡平性に対する潜在的脅威の要因

本質的に人間はバイアスを有しており、そのバイアスはAIシステムのライフサイクル全体（例：要件収集、モデル学習、モニタリングが性能不良、不正確な結果、または見逃されたシナリオを検出できない場合）を通じて導入される可能性があります。AI専門家および開発者も無意識のバイアスを持つ可能性があり、それが特定および対処されない場合、出力には限界が生じ、差別的となる可能性があります。さらにそのバイアス自体が認識されない場合があります。一方で、出力が正確であり、公平かつ衡平であったとしても、AIシステムに対する人間側の先入観により性能が不十分であると判断され、結果が拒否される可能性もあります。

### 8.3.1 不十分な学習および／または試験データセット

バイアスは主としてAIソリューションの開発および試験に使用されるデータに導入されそれがバイアスや差別を助長し、結果として有害な影響をもたらす可能性があります。AIが適用される集団を代表していない場合や、データセットが完全でない場合などのデータの制限により、不正確な結論が導かれる可能性があります。データの制限が透明化されていない、または認識されていない場合、代表されていない集団への不適切または意図しないAI適用が生じる可能性があります。データの頑健性や利用可能性の不足、例えば世界的に医療記録がデジタル化されていない状況は、十分にサービスが提供されていない集団の発生や、性能の低いモデルにつながる可能性があります。データの代表性が不十分な場合、利用可能なデータが対象集団と一致せず、代表性の低い集団への対応や、体系的なPV報告システムが十分でない地域のように整理されたデータが利用できない状況への是正が必要となります。そうでなければ、学習データに含まれる集団に偏ったシナリオが生じる可能性があり、学習データがすべての集団（例：すべての民族）を代表していない場合、不十分な学習データに基づくAIシステムは性能低下や、代表性の低い集団に対する差別につながる可能性があります。

データ不足は、データが存在しない場合だけでなく、利用可能な形式や構造で整理されていない場合、データの頑健性不足、またはすべての変数が十分に代表されていない場合にも生じ、モデル性能の低下、あるいはモデルの限界が認識されないことによる誤った結論を招く可能性があり、その結果として患者の健康転帰に悪影響を及ぼす可能性があります。例えば、主として高齢の非黒人男性を代表する臨床データを用いて急性腎障害（AKI）を検出するアルゴリズムを開発した場合、そのアルゴリズムは若年女性患者やデータに含まれていない民族に対して使用すると信頼性が低い可能性があります。<sup>4</sup>データ代表性の不均衡はデータの偏りを生じさせ、不均衡を増幅させる可能性があります。AIソリューションの出力を評価する際にバイアスの特定および評価が困難となる場合があります。

開発および試験に使用するデータセットが、想定される対象集団を完全に代表することが常に可能とは限らないことは認識されています。場合によっては、見かけ上類似しているデータセット間の差異が大きすぎて補完できないこともあります。一方で、適切な注意を払うことでデータセットを再利用できる場合もあります。この場合、開発者は適切な文書化を行い、不完全なデータに基づいて学習されたモデルの妥当性を示す必要があります。歴史的には、有害事象（AE）の既知の過少報告や刺激報告などのデータ制限により、PV活動においてバイアスが影響した事例が存在します。不十分なデータやデータの不均衡は、誤解を招く、または不正確な結果をもたらす可能性があります。変形性関節症の疼痛治療に使用されたCOX-2阻害薬ロフェコキシブ（Vioxx）は、販売中止後に

多数のAE報告が提出された刺激報告の重要な例です。<sup>5,6,7</sup> 安全性アラートが自発報告データベースにおける不均衡指標へ与える影響は、いわゆる注目バイアス (notoriety bias) の例を示しています。また、製造物責任を巡る集団訴訟などの法的手続きにより、法律事務所が原告候補を特定する過程で大量のAE報告が誘発される場合があります。<sup>10</sup> 依頼報告が自発報告を上回ることによってデータの不均衡が生じ、その結果、偏ったデータにより代表性の低い集団に関して誤った結論が導かれる場合、公正性と衡平性に対する脅威となり得ます。<sup>8</sup> 報告慣行、データ利用可能性、および変動性を考慮し、限界を理解するとともに潜在的バイアスを抑制する必要があります。これらのデータバイアスがAIソリューションに導入された場合、負の影響が増幅され、基礎となるバイアスの特定が困難なまま見過ごされる可能性があります。

### 8.3.2 十分にサービスが提供されていない集団

代表性の不足 (under-representation、過少代表性) は、十分にサービスが提供されていない集団セグメントを直接的に生み出し、サブ集団の特性が認識されない可能性があります。集団特異的な区分は、人口統計学的特性、疾患過程、遺伝的多様性、健康習慣の違い、ならびに医療レジメンや患者の期待に関する文化的差異などに基いて行われる場合があります。これらの差異はバイアスを導入し、負の影響または結果をもたらす可能性があります。これは、データが特定の集団に限定されている場合、排他的である場合、またはサブグループの特性が十分に理解されていない場合に生じ得ます。例えば、学習データセットにおいてアジアからの報告が十分に含まれておらず、かつ他地域と重要な点で異なっている場合、症例優先度判定アルゴリズムはアジアの特定の国からの報告に対して性能が低下する可能性があります。AIソリューションの開発および適用において、これらの集団を適切に考慮できるように、代表性の限界を特定および評価できる適切な課題領域の専門家 (subject matter experts : SMEs) が関与することが極めて重要です。

臨床試験においては、研究対象集団にサブグループが十分に含まれていない場合や、医療専門職や医療施設へのアクセスが限られているなど、より低い水準の医療しか受けられない場合、こうした潜在的な有害影響が研究者によって見逃される可能性があります。重要な情報を見逃さないために偽陰性を防ぐことへ重点が置かれる場合があります例えば特定のサブグループに限定または過剰に発生している潜在的有害性をPVプロセスが検出できないことが問題となり得ます。製造販売後において、特定の患者サブ集団で性能が低いAIソリューションが導入された場合、それらの集団からの有害事象 (AE) を検出できない可能性があります。一方で、重複検出においては偽陽性がより大きな問題となる場合があります、特定の国で重複疑いとして誤ってフラグ付けされる報告が多い場合その地域における安全性シグナルの検出遅延または見逃しにつながる可能性があります。

年齢関連集団 (小児、高齢者)、妊婦、ならびに希少疾患のような頻度が低いまたは報告が少ない事象、社会的スティグマを伴う事象など、しばしば十分に代表されない特別集団は、バイアス評価において考慮する必要があります。例えば、特別集団 (例：妊娠) に関するデータが限られた状態でシグナル検出支援を目的としたAIソリューションが導入された場合、誤解釈またはシグナルの見逃しによる負の影響が増幅される可能性があります。

対象集団を十分に代表していないデータに基づいて意思決定が行われる場合 (例：対象集団の代表性が限定的なデータに基づく製造販売後リスク最小化活動)、当該集団の患者安全を適切に確保できない可能性があります。AIソリューションにおいて不十分なデータを是正できない場合、非AIによるPV安全対策 (例：特別集団に対する強化されたモニタリング) が必要となる場合があります。

不利益を受ける可能性のある集団の詳細な特定、データセットに対して不均衡となる低頻度事象の特定、ならびに可能な場合に不十分なデータへ対応する包括的戦略の導入は十分にサービスが提供されていない集団に対する潜在的なバイアスおよび差別を低減することに寄与します。

### 8.3.3 人工知能ソリューションの設計

アルゴリズムは既存のバイアスや差別を助長してはならず、アルゴリズム設計自体が意図しない結果を生じさせる可能性があります。慢性疾患患者に対して積極的介入の恩恵を受ける可能性が高い患者を予測するモデルの開発にAIが使用された際、学習に用いられたデータセットが「必要性」ではなく「医療利用状況」に基づいていたため、黒人患者よりも白人患者に多くの資源が割り当てられる結果となりました。<sup>9</sup>医療システムおよび医療データ全体には国特有のバイアスが含まれている可能性が高く、AIを無批判に使用すると、これらのバイアスが予測結果に再現されます。AIの仕組みが不透明であること、能力に関する過度な期待、仕組みに対する理解不足、そして結論に疑問を呈する明確な手段が不十分であることにより、有害な結果が生じる可能性はさらに高まります。

人間が定義するパラメータやモデルによるデータ処理方法は、バイアスを導入したり不正確な結果を生じさせたりする可能性があります。AIソリューションの特徴量を個人の意識的または無意識的なバイアスに基づいて選択または設計した場合、出力結果は最適でない、あるいは誤ったものとなる可能性があります。生成AI (GenAI) のプロンプトエンジニアリング開発においては、プロンプト設計、具体性の不足、必要情報の欠落、または要求されるプロンプトの省略により、負のバイアスを含む出力が生成される可能性があります。個人の嗜好は意思決定に影響し、その結果としてデータ選択およびモデル開発にも影響を与えます。例えば、モデル開発者が自身に近い属性の集団に親和性を持つ場合（例：若年成人の開発者が小児または高齢者集団を十分に考慮しないデータを選択する場合）、このような状況が生じ得ます。

開発戦略には、バイアスを抑制し、代表的集団を考慮した完全かつ正確なデータ表現を達成するための、意識的かつ体系的なアプローチが必要です。戦略には、必要に応じたAIソリューションのレビューおよび調整、ならびに過去のバイアスの回避が含まれます。学習データおよびテストデータにおいて各集団がどのように表現されているかを文書化することは、限界やバイアスを把握し、軽減策の実装を支援するうえで有用です。対象集団を検討する際には、データが世界全体の集団を代表していることを確認し、バランス、人口統計学的均衡、および適切な分布と配分を確保する必要があります。

AIは医療分野において、安全性課題やコミュニケーションの特定に関連する一貫性、不一致、外れ値などのパターンや異常を検出する目的で、ますます活用されています。例えば、感情表現の一貫性を評価することで、人為的要因による不一致を特定し軽減することが可能となります。このような予防的アプローチは、不公平やバイアスのリスクを低減し、信頼性および客観性を向上させます。<sup>10</sup>

## 8.4. リスク、影響、緩和策

公正性と衡平性に対するAIの影響は、PVにおけるAIの適用方法、ユーザビリティ、性能ならびにプロセス内でAIが使用される位置に伴うリスクに依存します。データの制約および／またはアルゴリズム開発に起因してAIモデルに差別やバイアスが組み込まれている場合、その偏ったモデルの適用による負の影響は増幅されます。モデルは結果を増幅または歪曲し、不正確な結論、不適切な優遇または不利益、不平等、あるいは特定の集団や人口に対する差別を生じさせる可能性があります。

潜在的バイアスの影響を評価するためには、公正性を分析する手法を活用できます。AIシステムにおける公正性評価の現在の手法は、Li & Chignellが述べているように、規範的（価値基盤）、手続的（プロセス重視）、およびアルゴリズム的（技術的）アプローチに焦点を当てています。<sup>11</sup> 規範的アプローチは、AIシステム開発における理想的基準の達成を目的として、社会規範、共有価値、および原則に着目します。手続的アプローチでは、チェックリストや意思決定ツリーなどの定義された枠組みを用いた自己評価が可能となります。アルゴリズム的／統計的アプローチは、公平なアルゴリズム意思決定を支援する技術的手段に依存します。AIシステムの正式な影響評価には、公正性と衡平性に悪影響を及ぼす可能性のあるバイアスのスクリーニングを含める必要があります。このスクリーニングにより、潜在的リスク領域を特定し、負の結果を最小化するための対応的な軽減措置を講じることが可能となります。可能な場合には、詳細な評価の一環として適切な統計手法を用いてバイアス検証を行うことが、AIに関連する軽減戦略の策定を支援します。公正性に関する統計的検定の例としては、少数派／多数派に対するバイアス、単一症例バイアス、少数派または多数派への不適切な優遇を評価するt検定の使用が挙げられます。<sup>11</sup>

AIソリューションを導入前および導入後に評価し、顕在的または潜在的なバイアスを確認することは、リスク低減のための軽減措置につながります。AIソリューションの説明可能性は顕在的バイアスを明らかにする可能性があり、学習データの特性を理解することは、バイアスが導入され得る領域の把握、適切な使用範囲、ソリューションの限界、必要なヒューマン・オーバーサイトの程度、および期待性能の決定に役立ちます。バイアス評価を実施する際には、将来の再学習活動に向けた導入後データ注釈プロセスや、可能な場合の軽減戦略についても考慮する必要があります。

PVのシグナル検出活動において結果が欠落した場合、製品のベネフィット・リスクプロファイルが誤って表現される可能性があり、患者安全性に影響を及ぼす誤った人的判断または意思決定につながるおそれがあります。

異なるサブグループ間での性能に関する感度分析は、AIソリューションによって十分に対応されていない集団や人口を明らかにするうえで重要となります。網羅的な感度分析が実施困難な場合には、性能評価対象のサブグループ選定にリスクに基づくアプローチを適用する必要があります。これはサブグループまたは人口に関する学習データおよびテストデータの制約に依存する場合があります。

## 8.5. 主要な軽減戦略

- 各AIソリューションについて公正性と公平性の観点から評価を実施し、評価方法結果、および講じた軽減措置を明確に示します。
- AIモデル開発の各段階において共通して生じ得るバイアスを考慮し、AIモデルライフサイクルの各フェーズでバイアスに対処する主要な軽減戦略を策定します。<sup>4</sup> 学習データおよびテストデータセットが、すべての関連集団を十分に代表する完全なものであることを確保します。
- 可能な場合には感度分析を実施し、サブグループまたは人口を変更してAIモデルの結果を評価することで、公正性を検証し、期待される結果を確認するとともに十分に対応されていない集団を明らかにします。これは特に、AIソリューションの説明可能性が低い場合に重要です。例：
  - 性別およびジェンダー入力を変更し、出力への影響を評価する。
  - 「十分にサービスが提供されていない集団」の節で述べた、アジアなど特定地域において症例優先度判定アルゴリズムの性能が低下した事例を参照する。
- AIソリューションが説明可能であるにもかかわらず結果が期待どおりでない場合には、AIソリューションの設計、パラメータ、および特徴量選択についてバイアスの観点からレビューを実施します。
- 学習データの内容を透明性高く記述し、顕在的バイアスを明示するとともに、モデルの限界を明確化して不適切な適用や誤った結論を防止します。
- 開発およびモニタリング活動において必要な人間の関与の水準を決定し、正確な性能および公平な結果を確保するために必要な入力を提供します。

潜在的リスク領域の特定は困難ですが、バイアス、差別、および性能不十分なモデルを防止するうえで重要です。データ制約を完全に回避することは常に可能とは限らないため、データ特性の可視化を行うことで適切な適用およびリスク軽減の機会を確保できます。モデルの限界を理解し、それらの限界および潜在的バイアスをユーザコミュニティおよびAI性能を監視する関係者へ適切に伝達することが重要です。

## 第8章 – 参考文献

- 1 Hasanzadeh F, Josephson CB, Waters G, et al. Bias recognition and mitigation strategies in artificial intelligence healthcare applications. *npij Digit Med.* 2025;8:154. <https://doi.org/10.1038/s41746-025-01503-7> (Journal full text)
- 2 Government of Canada. Guide on the use of generative artificial intelligence. Ottawa: Treasury Board of Canada Secretariat; 2023. (Webpage accessed 21 March 2025)
- 3 Hamid AAA, Rahim R, Teo SP. Pharmacovigilance and its importance for primary health care professionals. *Korean J Fam Med.* 2022;Sep;43(5):290-295 <https://doi.org/10.4082/kjfm.21.0193> (Journal full text)
- 4 Nazer LH, Zatarah R, Waldrip S, Ke JXC, et al. Bias in artificial intelligence algorithms and recommendations for mitigation. *PLoS Digit Health.* 2023;2(6):e0000278. <https://doi.org/10.1371/journal.pdig.0000278> (Journal full text)
- 5 Pariente A, Gregoire F, Fourrier-Reglat A, Haramburu F, Moore N. Impact of safety alerts on measures of disproportionality in spontaneous reporting databases: the notoriety bias. *Drug Saf.* 2007;30(10):891-898. <https://doi.org/10.2165/00002018-200730100-00007>. (Journal full text)
- 6 Catalogue of Bias. Richards GC, Onakpoya IJ. Reporting biases. In: *Catalog Of Bias 2019*: [www.catalogueofbiases.org/reportingbiases](http://www.catalogueofbiases.org/reportingbiases) (Webpage accessed 22 September 2025)
- 7 Haguinet F, Bate A, Stegmann J-U. The futility of adverse drug event reporting systems for monitoring known safety issues: a case study of myocardial infarction with rofecoxib and other drugs. *Pharmacoepidemiol Drug Saf.* 2024;33(1):e5719. <https://doi.org/10.1002/pds.5719>. (Journal full text)

- 8 Jokinen J, Bertin D, Donzanti B, et al. Industry assessment of the contribution of patient support programs, market research programs, and social media to patient safety. *Ther Innov Regul Sci*. 2019;53:736-745. <https://doi.org/10.1177/2168479019877384> (Journal full text)
- 9 Obermeyer Z, et al. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*.2019;366:447-453. <https://doi:10.1126/science.aax2342> (Journal full text)
- 10 Bergman E, Sherwood K, Forslund M, Arlett P, Westman G. A natural language processing approach towards harmonisation of European medicinal product information. *PLoS ONE*. 2022;17(10):e0275386. <https://doi.org/10.1371/journal.pone.0275386> (Journal full text)
- 11 Li J, Chignell M. FMEA-AI: AI fairness impact assessment using failure mode and effects analysis. *AI Ethics*.2022;2:837-850 <https://doi.org/10.1007/s43681-022-00145-9> (Journal abstract )

## 第9章

# ガバナンスとアカウントビリティ

### ガバナンス – 原則

ガバナンスとは、医薬品安全性監視（PV）システムにおける人工知能（AI）の利用を統制および指揮するために用いられる人間の管理システムを指します。AIガバナンスフレームワークでは、AI指針原則への適合を確保するため、リスクマネジメントの実務および方針の実装が求められます。

### アカウントビリティ – 原則

アカウントビリティとは、PV義務を遂行するためにAIシステムを導入、運用および管理する組織および／または個人に対して、役割、責任および法的責任が明確に定義されることを指します。これには、規制当局、ベンダー、ユーザー、開発者、データ提供者、またはAIシステムの方針設定、開発、導入、維持および管理に関与する製薬企業を含む関係者が、適切なガバナンス措置を採用することが求められます。これにより、AIライフサイクル全体を通じて運用が想定された範囲内に維持されるとともに、予期しない結果にも対応できるようになります。

### 主要メッセージ

- ガバナンスでは、既存のPV品質マネジメントシステム（QMS）に、AIシステムの全ライフサイクル段階および影響を受けるプロセス全体を包含する包括的アプローチを組み込む必要があり、可能な限り早期に確立されるべきです。
- アカウントビリティは、PVにおいてAIソリューションを所有および運用する組織に帰属し、関与するステークホルダーの役割および責任を明確に定義する必要があります。AIシステム自体が責任主体となることはありません。
- システムおよびプロセスに加え、サービス提供者およびソフトウェアベンダーについても適格性評価を実施する必要があります。
- AIシステムがAI原則にどのように適合しているかについて定期的なレビューを実施し、継続的な規制遵守および性能維持を確保する必要があります。
- PVにおけるAIソリューションのガバナンスフレームワークグリッドは、AIシステムのライフサイクル全体を通じて主要要素を文書化するための構造化された指針として活用できます。
- ガバナンスおよびアカウントビリティは、AIシステムの事業上の利用目的や価値提案から独立して維持されるべきであり、これにより偏りのない意思決定が可能になります。

## 9.1. はじめに

前章までにおいて、AIシステムをPVシステム全体へ統合および実装する際には、リスクに基づくアプローチの採用、十分なヒューマン・オーバーサイトの提供、妥当性と頑健

性の実証、ならびに透明性、データプライバシー、公正性と衡平性への対応が重要であることを詳細に説明しました。本章では、AIを活用したPVにおけるガバナンスおよびアカウントビリティの指針原則について概説します。また、これら二つの原則の重要性、AIライフサイクルの中で特定のガバナンス対応を要する段階、さまざまなステークホルダーの役割と責任、規制上の監督、ならびに急速に進化するAI技術分野における継続的な研修の必要性についても論じます。

強固なガバナンスと明確なアカウントビリティは、AI導入の成功に不可欠です。これらの原則は、AIシステムが責任ありかつ倫理的な方法で使用され、規制要件に適合しながら、ステークホルダー間の信頼および透明性を促進することを確実にします。明確に定義された役割および責任により、すべてのステークホルダーが自らの義務を理解し、AIソリューションを効果的に監督できるようになります。

AI技術の進展に伴い、ガバナンスおよびアカウントビリティの枠組みも適応していく必要があります。新たなリスクおよび課題が出現するため、原則および実務の更新が求められます。これらの変化に対応し続けるためには、継続的なレビューおよび適応が不可欠です。これには、実務での利用を目的とした提案型ガバナンスフレームワークグリッドの改良も含まれます。

## 9.2. ガバナンスフレームワーク

PVにおけるAIソリューションのためのガバナンスフレームワークグリッド（以下、グリッド）（表9参照）は、PV用途として開発されたAIシステムのライフサイクル全体、すなわちコンセプト、開発、導入、監視の各段階にわたり、それぞれの原則に対応して対処すべき主要事項を特定するために設計された構造化ガイドです。

グリッドは、AIシステムの計画および監督のための構造化ガイドとして機能するだけでなく、自己評価にも活用できます。各アクションまたはプロセスをどの段階で実施すべきかを明確化することで、透明性、アカウントビリティ、リスクに基づくアプローチといった原則が一貫して遵守されることを支援し、AIのPVシステムへの統合を促進します。AI指針原則の遵守を確保することを目的として、ガバナンス組織が主要業績指標（KPI）を定期的にレビューすることにより、ギャップの特定およびAIソリューションの改善を促進できます。AIのPV利用に関する専門性と重点を持つガバナンス組織は初期段階では必要ですが、AIシステムが通常運用段階に入った後は、ガバナンスプロセスをPV全体の監督機構へ統合することを検討すべきです。AIシステムに重大な変更を要するリスクが発生した場合には、AI専任のガバナンス組織を再度関与させる必要が生じる場合があります。

グリッドの活用方法は複数あります。PV組織内の部門がAIベースの自動化を構想し、ベンダーまたは社内開発チームへ委託する際に、あらかじめガバナンス要件を定義する場合があります。あるいは、ベンダーが完成済みのAIシステムをPV組織へ提示し、本グリッドを適用してAI指針原則に照らした評価を行うことも可能です。AIソリューションの実装を成功させるためには、ガバナンス原則を早期に検討することが極めて重要です。これらの原則は、ベンダーシステムの開発または選定、導入、継続的管理、および廃止に至るまでの全過程を導くべきものです。初期計画では、潜在的リスクの特定および緩和戦略の策定に重点を置く必要があります。さらに、AIシステムが倫理的および規制上の基準と整合することに初期段階から焦点を当てることで、頑健かつコンプライアンスに適合したAIシステムの基盤を確立できます。

グリッドは、AIソリューションのライフサイクルにおける五つの段階で構成されています。すなわち、通常は業務部門が入力を提供する初期要件定義段階、続いて開発、導入前、導入後、通常運用の各段階です。これらの段階は、AIシステムの初期適格性評価

および反復の変更の双方に適用されます。また、AIソリューションのライフサイクルにおいては、特定の課題や新たな発見に対応するため、以前の段階へ戻る必要が生じる場合があることにも留意が必要です。各段階においてAI指針原則を考慮し、グリッドでは各原則が関連文書化のためのセルとして構成されています。特定のAIソリューションにグリッドを適用する場合、各セルには実施されたアクション、考慮事項、またはそれらが記録されている文書（SOP、作業手順書、ログファイル保管リポジトリ、レビュー済み性能指標、責任者またはレビュー組織名など）への参照情報を記載することが想定されています。各指針原則がライフサイクル段階全体でどのように適用されるかの例は以下に示されており、本グリッドの実務適用例は付録3（ユースケース）に示されています。

以下のライフサイクル段階の説明は、92～93ページに掲載されているガバナンスフレームワークグリッド（表9）に付随するものです。

**仕様および要求事項の収集**：本段階は、関係者を特定し参画させる初期段階であり、プロジェクトの目的、価値提案、適用範囲および機能を定義します。医薬品安全性監視（PV）専門家、データサイエンティスト、AI/機械学習エンジニア、ソフトウェアエンジニア、IT専門家、その他の領域専門家から構成される学際的チーム（第4章「ヒューマン・オーバーサイト」参照）は、通常、システム開発者、ソフトウェアベンダー、または社内IT開発チームによって管理されます。本段階は、開発者およびエンドユーザーに対するロードマップを提供し、開発プロセス全体の基盤を形成します。従来のソフトウェアと同様に、AIシステムの進化に伴い要求仕様の反復の見直しが必要となる場合があります、それに応じてグリッドも再検討される必要があります。

**開発および変更管理**：本段階では、学際的チームがAIシステムの取得、構築、または改修に注力し、必要な機能を備え、かつガバナンス原則に準拠して構築されていることを確保します。AIシステムを新規開発する場合であっても、ベンダーシステムを選定する場合であっても、これらの原則は全過程にわたり適用されます。

**導入前および変更後の「承認（sign-off）」**：本段階では、AIシステムは開発段階からPVプロセスへの導入段階へ移行します。実装前には、AI特有の観点に基づく出力レビューを含む包括的なバリデーションおよび承認プロセスを実施し、AIシステムまたはその変更が導入可能な状態にあることを確認する必要があります。通常、この段階ではPV専門家がAIシステムによって生成された結果および当該システムが使用される業務プロセスの適応について責任を負います。本段階の文書化には、リスク評価、原則遵守状況のレビュー、承認書式、バリデーション報告書、ならびに承認手順や導入により影響を受けるPVプロセスを規定するSOPへの参照などが含まれる場合があります。

**導入後および変更後の「ハイパーケア」**：導入後、本段階はAIシステムの性能または最新変更の影響を即時に監視するうえで極めて重要です。PVプロセスにおける実運用では誤った前提、設計上の欠陥、または意図しないバイアスなど、前段階では顕在化しなかった問題が表面化する可能性があるため、予期しない問題を迅速に特定し解決するための集中的観察期間となります。本段階は従来ソフトウェアのハイパーケアとは異なり、AIシステムでは即時修正が困難な場合があります、人間による介入やヒューマン・オーバーサイトの強化などの措置が必要となる場合があります。文書化が求められ、観察期間中の最新変更に関するインシデントログや性能分析報告書などが含まれる場合があります。

**通常運用**：本段階は、AIシステムがPVプロセスへ完全に統合された状態を示します。継続的な監視、保守、および文書化を実施し、十分な監督を確保するとともに、事前に定義された主要業績指標（KPI）の監視を通じて傾向の把握を可能にします。本段階では定期報告書、継続的対応のログ、ならびに当該レビュー手順を管理するSOPまたは作業手順書への参照が含まれ、モデルが完全運用状態にあることを反映します。

なお、導入後段階または通常運用段階における新たな発見により、AIシステムの運用を一時停止し、改善のため導入前段階へ戻す必要が生じる場合があります。

表 9：ガバナンスフレームワークのグリッド

出典：CIOMS Working Group XIV

	仕様収集・要求定義	開発および変更管理	導入前サインオフ	導入後ハイパーケア	定常運用
リスクに基づくアプローチ	リスク評価 (理論的評価) ▶ AI システム ▶ 使用状況 (Context of use) ▶ リスクの影響および発生可能性 リスク低減計画	リスク評価 (実証的評価) 監督戦略の定義 変更管理 (職員教育計画を含む)	性能評価に基づくリスク低減計画の調整 ヒューマン・オーバーサイト戦略 職員の展開	リスク評価 (実証的評価) のための集中的または重点的モニタリング (高リスク領域を対象) 必要に応じた低減措置	定常モニタリング (例：モデルドリフトリスク) 必要に応じた低減措置
ヒューマン・オーバーサイト	学際的専門性の確保 ヒューマン・オーバーサイト戦略の検討	監督戦略の定義 変更管理 (職員教育計画を含む)	ヒューマン・オーバーサイト戦略の微調整 職員教育の展開	監督戦略の実装 集中的または重点的介入	定常的監督活動 必要に応じた監督戦略の調整および最適化
妥当性と頑健性	ユースケースおよび適用領域の仕様定義 参照標準 (Reference standard) の仕様定義 ベンチマークの仕様定義 再現性要件	モデル学習およびバリデーション参照標準の開発または取得	性能評価 ベンチマーク比較 システム試験の正式実施	性能モニタリング 継続的インテグレーションおよびデプロイ	定期的性能モニタリング
透明性	[組織 → 開発者] モデル要件 ▶ 想定用途 (Intended use) ▶ 人間-コンピュータ相互作用 ▶ 説明可能性 ▶ 期待される出力 性能評価要件 ▶ 範囲 (Scope) ▶ 参照標準	[開発者 → 組織] モデル情報 ▶ アーキテクチャ ▶ パラメータ ▶ 許容入力 ▶ 期待出力 ▶ 標準AIコンポーネント ▶ モデル学習および検証 ▶ 既知の限界 モデル開発、デバッグおよび文書化を支援する説明可能性	【開発者 → 組織】 性能評価 ▶ 範囲 ▶ サンプルング ▶ 参照標準 ▶ 人間による入力 ▶ 要約指標 ▶ ベンチマーク ▶ サブセットおよび感度分析 ▶ 定性的レビュー 妥当性と頑健性、公正性と公平性評価を支援する説明可能性	[開発者 → 組織、および組織 → エンドユーザー] 性能評価 ▶ 逸脱事項 妥当性と頑健性、公正性と公平性評価を支援する説明可能性	[組織 → エンドユーザー (および規制当局)] AI使用の開示 エンドユーザーとの信頼構築を支援する説明可能性

データ プライバシー	仕様収集・要求定義	開発および変更管理	導入前サイネオフ	導入後のハイパーケア	定常運用
	利用目的の仕様定義 ▶ データソースの仕様定義 ▶ 識別子を含むデータ要素の特定 ▶ データの管轄区域/来歴 (provenance) の確認 ▶ プライバシー・バイ・デザイン (データ最小化) ▶ データ保護影響評価 (Data Protection Impact Assessment)	学習データセット選定、アルゴリズム設計  学習データセット選定、アルゴリズム設計および認知バイアスへの配慮 ▶ 以下の回避 ▶ 明示的または潜在的な不公平バイアス ▶ 不十分なデータ包含	テストセット (公開予定の場合、例：公開ベンチマークとして使用する場合は、地域の法令/規制/ガイドランスに整合したデータプライバシー確保が必要)  導入前性能評価 ▶ 参照データセットの不備 (例：未整備、不十分な代表性) ▶ アルゴリズム設計 ▶ 人的/認知バイアス	完全データセット/継続的に蓄積されるデータセットでモデルを運用する際のデータプライバシー配慮の遵守 ハイパーケア期間中は逸脱への注意強化	定常運用におけるデータプライバシー問題を特定し是正する継続的プロセス
公正性と公平性	利用状況の定義 — 許容される適用範囲	導入を支援するKPI合意 主要役割 (例示) ▶ PV専門家 ▶ データサイエンティスト ▶ AI専門家 ▶ IT専門家 ▶ 倫理専門家	性能評価に基づくKPI精緻化 主要役割 (例示) ▶ PV専門家 ▶ データサイエンティスト ▶ AI専門家 ▶ 上級管理職	リスク低減戦略の承認 主要役割 (例示) ▶ PV専門家 ▶ AI専門家 ▶ IT専門家 ▶ データ保護責任者 ▶ サイバーセキュリティ専門家	定常モニタリング モデルシフトや不十分な学習データ (代表性不足集団、特別集団) に起因する性能低下
ガバナンスと アカウントアビリティ	AIシステムが既存PVシステムへどのように適合するかの検討 主要役割 (例示、網羅的ではない) ▶ PV専門家 ▶ AI専門家	導入を支援するKPI合意 主要役割 (例示) ▶ PV専門家 ▶ データサイエンティスト ▶ AI専門家 ▶ IT専門家 ▶ 倫理専門家	性能評価に基づくKPI精緻化 主要役割 (例示) ▶ PV専門家 ▶ データサイエンティスト ▶ AI専門家 ▶ 上級管理職	リスク低減戦略の承認 主要役割 (例示) ▶ PV専門家 ▶ AI専門家 ▶ IT専門家 ▶ データ保護責任者 ▶ サイバーセキュリティ専門家	KPIの監視を含むPV品質マネジメントシステム全体への統合 主要役割 (例示) ▶ PV専門家 ▶ AI専門家

以下に示す非網羅的な例は、ガバナンスフレームワークグリッドにおける各ライフサイクル段階に関連して、各AI指針原則に関して考慮すべき事項を示しています。

**透明性：**開発段階では、変更理由およびモデル学習に使用されたデータを含め、開発活動に関する包括的な文書化の作成に重点が置かれます。AIシステムおよび学習データの記述が透明であることを確保し、不適切な適用や誤った結論を防ぐために限界事項を明確に示す努力が求められます。導入前段階では、モデル性能評価および公正性・公平性に関する実証的証拠を追加することで透明性がさらに強化されます。また、作成された文書は、異なるステークホルダー間で想定用途に対する理解が一貫することを保証する必要があります。定常運用段階では、最も重要な透明性はエンドユーザーおよび継続的な性能評価とモニタリングを担う者に対して確保されることになります。

**説明責任：**すべての段階を通じて、IT部門、ベンダー、またはPV専門家などに対して責任を割り当て、文書化することが一貫して必要となります。これにより、AIシステムの開発、変更管理、導入、および性能について、どの時点でも誰が責任を負うのかが明確になります。

**リスクに基づくアプローチとヒューマン・オーバーサイト：**これはAIシステム開発に関連するリスク水準の特定から開始されます。必要に応じて、人間の専門家による頑健な手法開発および性能評価を確保するため、明確なアノテーションガイドラインの策定を含む場合があります。AIシステム性能を評価する際には、低頻度環境に対する特別な配慮が必要です（第5章「妥当性と頑健性」における性能評価の議論も参照）。次の段階として、「human-in-the-loop」の定義やその他の監督手段の設定、さらにはユーザーインターフェイスにおけるリスク低減要件の構築など、適切な低減戦略を提案します。その後、導入前段階でヒューマン・オーバーサイトを再定義し、実運用に基づく観察結果を踏まえて定常運用段階において定期的にこれらの概念をさらに洗練させていきます。この段階的アプローチは、AIシステムがライフサイクルを通じて進展するにつれて、リスク管理も進化させる必要があることを示しています。一般に、AI指針原則への適合を確保するためのすべての措置において、リスクに基づくアプローチの採用が推奨されます。

AIシステムへのいかなる変更も、初回導入時と同様に厳格なガバナンス検討を受けなければなりません。これにより、変更がシステムの完全性または性能を損なわないことを保証します。透明性および説明責任を維持するためには、文書化およびバリデーションが不可欠です。更新および改修を適切に管理し、ハイパーケアに基づいて定常モニタリングを超えた性能および品質確認を行う導入後段階を考慮した変更管理プロセスを整備する必要があります。同時にコンピュータ化システムバリデーション要件も満たす必要があるため、AIシステムのバージョン管理は、他のソフトウェアのバージョン管理から分離して管理することが望ましいとされています。

## 9.2.1 ガバナンス体制および説明責任の割当て

各原則に対するレビューを効果的に実施し、対応策およびリスク評価について合意形成を行うためには、ガバナンス体制（governance body）を構築することが望ましいとされます。このグループは理想的には、多様性を備えた部門横断型チームであり、エンドツーエンドのプロセスおよびその中における自動化の範囲について十分な理解を有している必要があります。すべての関連ステークホルダーからの代表者を含めるべきであり必要に応じてソフトウェアベンダー代表の参加も検討されます。このような多様性および職務分離により、AIシステムに対する広範かつ均衡の取れたレビューが確保されます。

ガバナンス体制は、AIシステムの開発、導入、性能、および継続的管理を監督し、すべての活動がAI指針原則および規制基準と整合していることを確保します。また、ガバナ

ンス体制は各ライフサイクル段階における責任者を決定し、PVプロセスへAIシステムを導入する前に必要な文書の承認（サインオフ）を含みます。AI導入の推進要因としてビジネスケースが存在するケースが多いため、ライフサイクル各段階におけるガバナンス措置に十分なリソースを確保する観点から、該当プロジェクトマネージャーまたはスポンサーもガバナンス体制に含めるべきです。

従来型ソフトウェアとは異なり、AIシステムにおけるガバナンス体制は、定義された間隔で、必要に応じてアドホックにもAI利用指針原則への適合状況をレビューする必要があります。これは、AI分野が急速に進展していること、および入力、ルール、その他予期しない要因の変化がAIシステムへさまざまな程度の影響を与え、場合によっては重大な影響を及ぼし得るという固有のリスクによるものです。導入済みAIシステムの再評価の適切な頻度および範囲は評価される必要があります。また、必要に応じて介入またはAIシステムの停止を行うための措置を準備しておくべきです。

ガバナンス体制がAIシステムが定常運用段階に到達したと判断した場合、ガバナンスはプロセスオーナーへ引き継がれ、PVシステム全体のモニタリングプロセスへ統合されます。ただし、AIシステムに重大な改修または停止を要するリスクが発生した場合には、AI専任のガバナンス体制を再度招集する必要があります。また、ガバナンスフレームワークグリッドに対するバージョン管理の導入も検討すべきです。

### 9.3. トレーサビリティとバージョン管理

トレーサビリティおよびバージョン管理は、AIシステムを管理する上で極めて重要な要素です。特にPVのような規制領域では、誤りが患者安全または公衆衛生へ影響を及ぼす可能性があるため、その重要性はさらに高まります。これらはAIシステムの過去バージョンの評価および再現性を可能にし、監査目的においてもしばしば要求されます（第5章「妥当性と頑健性」における確率的要素を含むAIシステムの議論も参照）。既存のバージョン管理フレームワークにおける一般的なベストプラクティスは、AIシステムのバージョン管理において有用な指針となり得ます。これらはエンドツーエンドプロセスに関する他の関連システムと併せて文書化されるべきです。また、ユーザー受入試験（UAT）環境および本番環境の双方において、明確な変更管理プロセスを含む必要があります。

AIシステムの文書化は、そのライフサイクル全体を対象とすべきであり、正当化、初期スコopingおよび構想段階、開発、導入、バリデーション、導入後管理、ならびに廃止までを含む場合があります。これにより、事前に定義された計画から逸脱した場合の正当化や判断理由を含め、重要な手順および意思決定の再現および追跡が可能となります。従来のコンピュータ化システムバリデーションと同様に、開発環境内または開発前に実施されたすべての実験を逐一記録する必要はありません。しかし、その実験または分析結果がAIシステムの評価または導入方法に影響を与える場合には、その判断根拠を文書化する必要があります。開発段階の結果や知見に基づいて意思決定が行われた場合も同様に文書化されるべきです。

開発段階において、AIシステムは継続的な実験および反復的改善を経ます。開発チームとPV組織との間の透明性は、効率性の確保およびシステムが目的適合性（fit-for-purpose）を満たすことを保証するために極めて重要です。開発者は複数のモデルバージョンを作成し、さまざまな機能を試験し、異なるトレーニングデータセットで実験を行う場合があります。このような状況では、モデルアーキテクチャの大幅な変更、新たなデータセットの導入、または性能指標の重要な変化といった主要なマイルストーンを明確に記録することに重点を置くべきです。これにより、開発者は多数の小規模な調整や実験に埋もれることなく、モデル進化の過程および主要変更の影響を把握できます。

AIシステムが開発段階から本番環境における定常運用へ移行すると、厳格なトレーサビリティおよびバージョン管理の必要性は大幅に高まります。本番運用されるモデルの各バージョンは詳細に文書化されるべきです。各バージョンのソースコードに加え、基盤となるモデルアーキテクチャ、トレーニングおよびテストデータセット、性能評価結果も文書化する必要があります。規制の観点からは、EUにおいてはPV System Master File (PSMF) などの文書内にAI構成要素を記載することが適切とされています。

導入後におけるAIシステムの継続的改善および適応についても文書化が必要です。これは人間の専門家によって開始される場合もあれば、性能劣化やモデルドリフトの事前定義されたモニタリングを含め、AIシステムの運用設計自体に組み込まれている場合もあります。Software as Medical Deviceに関連するこれらの課題の一部についてはFDAにより説明されています。<sup>1</sup>

外部AIコンポーネント（事前学習モデルやライブラリなど）を統合する場合には、それらのバージョンを文書化することが重要です。特にそれらがモデル性能に重要な役割を果たす場合は必須となります。ただし、すべての反復ごとに記録する必要はなく、重要なマイルストーン時点で文書化することで十分な場合もあります。例えば、AIベースの静的システムについては、AIシステムライフサイクル各段階における文書化の考慮事項を含む特定の文書化アプローチが既に提案されています。

### 9.3.1 強化型医薬品安全性監視システムにおける役割および責任

組織は、自社のPVシステムに関連する品質プロセスについて責任を負い、その中にはシステムオーナーによるAIコンポーネントの監督も含まれます。監督活動は、適切な監督のもとで第三者によって実施される場合もあります。EU AI法などの規制では、AIリテラシーの促進や公正性と衡平性の確保を支援するための特定の役割の設置が組織に求められる場合があります。AIシステム自体が責任主体となることはできません。AIの安全かつ責任ある利用を確保するためには、ヒューマン・オーバーサイトが不可欠です。AIの取り組みに関与するすべてのステークホルダーについて、明確な役割と責任を定義する必要があります。

AIの導入により、PV専門家の役割は変化しつつあります。すでにAIは、AIシステムの監督やその出力結果の解釈といった新たな業務を生み出しています。PV専門家はこれらの変化に適応し、自らの責務を果たすために新たな技能および能力を習得する必要があります（第4章「ヒューマン・オーバーサイト」参照）。これは特に、ガバナンス組織のメンバーやライフサイクル各段階において責任者として指名された者にとって重要です。ガバナンスフレームワークグリッドは、特定の工程において新たな活動が必要となるかを関係者が評価できるようにし、早期に教育・訓練ニーズを明確化することを可能にします。

従来のソフトウェア提供者との関係と同様に、AIシステムベンダーとPV専門家との協働は極めて重要です。この協働により、AIシステムがPV要件およびガバナンス原則を満たしていることを確保できます。ベンダーおよびAIシステムの定期的な監査および適格性評価、継続的な性能モニタリング、事業継続計画は、コンプライアンス維持および開発水準の確保に不可欠です。効果的な協働と監査は透明性および説明責任を強化し、AIシステムの信頼性確保、規制要求への適合、ならびに査察対応可能な状態の維持につながります。

規制当局もまた、PVにおけるAIの監視に重要な役割を果たします。査察を通じて、AIシステムが規制基準およびガバナンス原則に適合しているかを監督します。また、規制当局はPVを含む医薬品ライフサイクル全体におけるAI利用に関するガイダンスの策定を進めています（第2章「ランドスケープ分析」参照）。AIシステムをPVシステムへ統合するには第6章「透明性」で述べられているように、PSMFなど適切な規制文書への反映が必要となります。

今後、PV査察ではAIシステムへの注目がさらに高まる可能性があり、査察官はAI関連文書、性能指標、およびガバナンス実務を確認することになります。これらを適切に評価するためには、査察官自身にもAIおよびデータサイエンスに関する技術的知識が求められます。そのため、新しく急速に進展する分野に対応するための継続的な能力開発および教育が必要となります。

AI活用の成功には、イノベーションの推進と規制遵守および指針原則への適合とのバランスが重要です。これは、責任あるイノベーション文化の醸成を意味します。これらの目標は、AIシステムのKPIを定期的にレビューする仕組みを含む効果的なガバナンスプロセスの構築によって達成できます。

## 第9章 – 参考文献

- 1 U.S. Food and Drug Administration (FDA), Health Canada, Medicines and Healthcare products Regulatory Agency (MHRA). Good machine learning practice for medical device development: guiding principles. Silver Spring (MD):U.S. Food and Drug Administration; 2021. (Full text accessed 21 March 2025)

# 第10章

## 医薬品安全性監視における 人工知能の開発と展開に関する 今後の検討

### 10.1. 医薬品安全性監視における人工知能の進化と将来

本章では、現在の適用状況から将来において人工知能（AI）が医薬品安全性監視（PV）へどのような影響を及ぼし得るかという展望に至るまで、AIがPVに変革をもたらし続けている影響について考察します。本報告書の前章におけるCIOMS Working Group XIVの議論は、共通の原則に基づいています。ユースケース（付録3参照）では、評価中または導入段階にあるさまざまなAIシステムを詳細に示し、PV分野における有効性の評価を提供しています。将来を予測しようとする際には、AIの発展の軌道がダイナミックであり、かつ高度に予測困難であることを認識することが不可欠です。実際に確実に予測できる点は、AIが医学分野において広範に導入され多くの側面、場合によっては創薬および医療実践のほぼすべての領域を、基礎研究から臨床現場に至るまで、さらにはPVにおいても革新する可能性があるという点のみです。このため、本章は本報告書の前章で述べられたAIのさらなる発展を基盤とし、PVにおけるAI利用がより普及し高度化するにつれて、原則に基づく応用がどのように進化していく必要があるかを展望します。

本章では、規制当局、医療従事者（HCP）、ならびにその他の産業関係者を含むPVステークホルダーに向けて、PVにおけるAIの安全かつ衡平な導入を確保するための考慮事項を提示します。現在PV専門家に求められている技能は、将来必要とされる技能とは異なる可能性が高く、PVにおけるAIの設計、開発、導入、および日常運用への関与が求められるようになります。本章の例示は、PVにおけるAI導入の方向性と巨大な潜在力を示すものですが、一定程度の推測を含むものであり、網羅的なものではありません。AIはPVを革新する可能性を有するだけでなく、PVの従来の境界を解消し、医学領域全般における適用範囲をさらに広げていくと考えられます。

現在に至る10年間は、PVにおけるAI導入の初期段階に位置付けられますが、AI分野、特にGenAIが急速に進展していることを認識することが重要です。AIのさらなる広範な導入は、今後数年間においてPV戦略の考え方やアプローチの変化を必要とする可能性がありPVという分野を従来の枠組みを超えて発展させ、AI主導型医学の進化と整合する形で安全性データの自己検出型リアルタイム監視へと変革していく可能性があります。例えば、大量の安全性データを迅速に解析・抽出し、症例報告やシグナル検出に活用する能力が挙げられます。AIのこの能力を活用することで、PVは報告および評価を中心とした反応型の活動から、予測および予防を中心とした将来志向型のアプローチ、さらにはリアルタイムまたは準リアルタイム学習システムへと進化していきます。

進化の初期段階はすでにPVの中核業務である症例管理および安全性監視に影響を与え始めており、AI強化型医療の進展により、報告および評価中心の活動から予防志向へと移行しつつあります。これらの技術は、優先度の高い事象への対応時間短縮、大量かつ多様な安全性データ（文献を含む）の解析能力向上、症例報告の自動生成、シグナル検出の実施などを通じて、PV専門家の手作業負担および業務負荷を軽減する可能性があります。<sup>1,2,3,4,5</sup>

## 10.2. 医薬品安全性監視の長期的・将来的な変革的役割：予測から検出・予防へ

高度なAIシステムは、現在のPVの境界を超える可能性を有しており、例えば、自動化または拡張された意思決定、AIエージェント、ならびにAIの学習、モデリングおよびシミュレーションを改善し、個別化医療を最適化するための量子コンピューティングに支援された業務へと拡張されることが想定されます。<sup>6,7,8,9</sup> 近似推論能力を備えたAIは、曖昧性や部分的な真理値を取り扱うことが可能であり、例えばソーシャルメディア投稿から得られる安全性データや、複数システムに分散した断片的な安全性関連データの評価に対応できる可能性があります。このようなAIシステムは、症例分類（例：因果関係の割り当て）やその他の医学的判断を要するPV業務において、PV専門家がより精緻な意思決定を行うことを可能にします。これは、データが不完全または相互に矛盾している症例において特に有用であり、グレーゾーンにおいて高度で状況認識的な分析および／または医学的判断が求められる場合に役立ちます。<sup>10,11</sup> 本報告書の前章ではAIモデルの現在の限界について言及していますが、AI基盤が継続的に進化していることをより明確に認識することが推奨されます。AIシステムの進化により、現在のモデルが有する制約の一部、あるいはそのすべてが将来的に改善・解決されることが期待されています。

将来的には、PV専用設計された専門家AIシステムが、当該分野において豊富な経験を有する専門家または専門組織の判断および意思決定プロセスを模倣する可能性があります。これらのシステムは人間の専門家を置き換えるものではなく、その能力を拡張し、より精緻かつ効率的な意思決定を可能にするものと考えられます。理想的なPV専門AIシステムは、腫瘍学、免疫学、ワクチン、医療機器といった治療領域、さらにデジタル治療など大きく異なる治療オプションに特化した高度な分析精度を組み込み、それぞれの治療、疾患、および患者集団の複雑性に適応しながらPV業務を実施または支援できるよう設計されます。このようなシステムの開発には多大な投資が必要ですが、次世代のターゲット型PVソリューションを推進する潜在力を有しており、患者安全性向上における重要なイノベーションとなる位置付けが期待されます。

将来的には、従来のPVが主として有害事象の検出および処理を行う活動から、技術主導型の最前線分野へと移行し、「自己 (self)」(人間または臓器：心臓、腎臓、肝臓、肺など)と情報を検出・評価・共有できる技術を設計する分野へ変化している可能性があります。<sup>12,13,14,15</sup>

これにより、医療従事者および患者が有害症状の発現前に是正措置を講じることで、監視および予防により積極的な役割を担うことが可能になります。AIを活用するPVは、異常を能動的に評価し、自己報告および自己学習を通じて将来の異常発生を防止し、患者の健康と安全の維持を継続的に促進する技術を開発する機能へと進化していくと考えられます。これには、真にAIによって実現される能動的かつ自己制御型の監視およびリスク低減が含まれます。

## 10.3. AIの今後の開発と展開および指針原則

CIOMS Working Group XIV のメンバーは、本報告書をAIのPV利用に関する共通原則を中心として構成するという慎重な決定を行いました。これは、この変革的技術が指数関数的成長の段階にあるという認識に基づくものでもあります。現時点の事例に過度に依存した規定的な報告書は、特に他の医療分野のAI技術が活用された場合、急速に陳腐化する可能性があります。著者らは、PVにおけるAI利用の共通原則は、予見可能な将来にわたり持続性を有すると考えています。一方で、指針原則（guiding principles）がどのように適用されるかについては不確実性が残ります。これらの原則自体は頑健であり継続すると期待されますがAIの技術的進展および医療を受ける個人による検出、予防、意思決定へのAI活用と並行して進化していく可能性が高いと考えられます。以下では、それぞれの指針原則に関する潜在的な影響について考察します。

### リスクに基づくアプローチ

第3章では、リスク低減策を含むリスクに基づくアプローチおよび必要な規制枠組みについて論じています。

AIの普及および進展により、継続的自己学習および潜在的に自律的なAIシステムの実現が可能となり、PVの大きな進歩および患者や医療従事者への利益がもたらされる可能性があります。

しかしながら、このようなシステムには懸念およびリスクも伴います。例えば重要な懸念として、AIがリアルワールドの診療環境における医薬品のベネフィット・リスクプロファイルに対する理解を歪める可能性が挙げられます。従来、これらのプロファイルは自発報告システムおよび計画された調査研究を含む慎重に設計された枠組みによって評価されてきました。しかしAI主導型システムは、例えば高リスク患者に対するAI強化PVシステムへのアクセス制限や適応外使用の抑制などを通じて、意図せず処方行動を制限する可能性があります。

さらに問題を複雑にする要因として、このようなシステムの導入状況および利用可能性が医療制度や地域によって異なる可能性があり、解釈が困難なデータパターンの不一致を生じさせます。このような分断された状況は、AIシステムが処方意思決定に与える真の影響を不明瞭にし、患者アウトカムへの実際の影響評価を困難にします。加えて、AIの誤った解釈や不適切な利用は患者安全性を著しく損なう可能性があります。ヒューマンファクターおよび人間工学（Human Factors and Ergonomics：HFE）の原則は、AI設計の簡素化を支援し、人間のパフォーマンスを最適化することでAI結果の理解向上に寄与します。HFEは、人間とシステムの他要素との相互作用を理解し、人間の福祉およびシステム全体の性能を最適化することを目的とした科学分野であり、原則、データ、および方法を用いてシステム、製品、環境の設計および改善を行います。<sup>16</sup>

このような高度AIシステムの監督およびリスク低減には、動的なリスク評価フレームワークが求められます。すなわち、ほぼリアルタイムのモニタリングと適応的評価プロセスを統合した枠組みです。これらの進化するリスクを、患者を含むすべてのステークホルダーに対して効果的に伝達することが極めて重要です。リスク低減の一環として、医療分野のリーダーはAIの進化する特性を考慮した柔軟なガバナンスモデルを採用し、透明性、説明責任、および衡平なアクセスが常に重視されるようにする必要があります。

リスクに基づくアプローチに関する第3章で述べたとおり、リスク低減では希少かつ予測されない異常事象（ブラックスワン事象）の特定も考慮する必要があります。現在のPVシステムは一般的な安全性リスクの予測、評価、および管理には十分対応可能ですが、特に高度AIシステムが導入される場合には、このような外れ値事象の検出にも適応していく必要があります。

## ヒューマン・オーバーサイト

第4章では、AI利用がますます活動に組み込まれ普遍的となる中でのPVにおける従来の役割の変化および変革を含め、ヒューマン・オーバーサイトについて説明しています。

AIシステムがより広範に普及し自律性を高めるにつれて、ヒューマン・オーバーサイトの役割は必然的に変化します。Human-in-the-Loop（HITL）アプローチの維持は引き続き重要である可能性が高いものの、PVの一部を含む高度に複雑または高リスクの用途では十分ではない可能性があります。一方で、AIシステムがデータレビューや自己プロセス制御において人間の能力を上回る場合には、ヒューマン・オーバーサイトの役割は実質的に変化し、その重要性が低下する場面も生じ得ます。<sup>17,18,19</sup>

この進化する環境では、PV専門家は新たなスキルセット、AIシステムのベンチマーキングツールおよび指標を習得し、AI主導型システムを効果的に監督するための専門的訓練を受ける必要があります。これには、AI支援テストおよびベンチマーキング、KPI、ワークフロー改善、ドリフトおよびバイアスマニタリングなどが含まれますが、これらに限定されるものではありません。監督の焦点は従来型の方法を超え、AIの挙動を理解し解釈し、方向付ける能力へと拡張される必要があります。さらに、PVにおいてAIシステムを監督する専門家および技術者に対して、適切なモニタリングを確保するための動的かつ継続的な教育訓練の実施が必要です。しかしながら、特定の高度に複雑または極めて高リスクな用途では、HITLによる監督のみでは不十分となる可能性があります。そのような場合には、十分なヒューマン・オーバーサイトを維持するために、新たなツールの開発と訓練を組み合わせる必要があります。これらの能力および補完的監督手段を育成することで、PV専門家は患者安全および公衆衛生を守るうえで、ヒューマン・オーバーサイトを引き続き実効性あるものとして維持できます。

## 妥当性と頑健性

第5章では、妥当性と頑健性について、多分野連携の必要性、参照標準、およびAIシステムの妥当性と信頼性を確保するために必要となる性能評価について論じています。

AIがより深く組み込まれ高度化するにつれ、その進展と並行してデータ完全性を検証し保証するための適切な方法およびシステムを構築することが課題となります。例えば、リアルタイムまたは準リアルタイムで膨大なデータを処理できる可能性がある場合、誤シグナルのリスクを回避するために拡張可能なバリデーション手法が必要となります。そのため、PV担当者には新たなスキルの習得、さらにはAI対応システムの検証要求に対応するための新しい科学分野の形成や、分野横断的かつ多職種から構成されるPVチームの構築が求められる可能性があります。第5章で述べたとおり、生成AIを含むAIの批判的評価に関するベストプラクティスは現在も進化途上にあり、今後より理解が進み、一貫して活用されるようになると考えられます。高度技術と組み合わせたAI利用では、新たな標準およびバリデーション手法の確立が必要となり、例えば埋め込み型チップを含む神経技術、ナノテクノロジー、スマート臓器（smart organs）などによって生成されるリアルタイム／準リアルタイム安全性データの最適化を目的とした導入方法についても検討が必要となります。

## 透明性

第6章では、AIシステムの透明性および説明可能性、ならびにそれに関連する課題について説明しています。

AIがますます広範に普及するにつれて、その導入状況を追跡し意思決定プロセスを理解する能力は低下する可能性があり、説明可能性および透明性に重大な課題をもたらします。AIシステムは複雑な統計的プロセスや高度なプログラミング、あるいはAI自身によって生成されたプログラムを反映する可能性があります。その結果、完全な透明性の必要性、さらには実現可能性自体が新たな課題に直面する可能性があります。AIシステムへの信頼が高まり、事前に定義された信頼水準を満たすようになるにつれて、透明性に対する期待も進化する必要があります。

現在のデータ解析、統計解析、およびシグナル検出におけるAIの役割と同様に、AIが下流の意思決定へ与える正確な影響を追跡することは、今後ますます困難になる可能性があります。経験ベイズ幾何平均 (Empirical Bayes Geometric Mean : EBGM) による不均衡分析モデルにおける事前分布の複雑性が広く受け入れられている一方で詳細に検証されることが少ないのと同様に、AI生成アウトプットへの確立された信頼が形成されることで、誤りや計算不備が発生した場合には適切な是正措置が取られるという前提のもと、関心の焦点が変化する可能性があります。

同時に、AIへの信頼が確立されるにつれて、説明可能性に対する重点も同様に変化する可能性があります。透明性は引き続き重要であるものの、その最も重要な価値はインシデントやエラー発生時に顕在化する可能性があります。航空分野におけるフライトデータレコーダーの役割と同様に、説明可能性は常時要求されるものというよりも、障害の理解およびシステム改善を目的として重要となります。

この変化はPVの意思決定に大きな影響を与え、迅速な介入および準リアルタイムでの根本原因分析を重視する方向へ導く可能性があります。将来的には、組織はAI性能向上による利益と要求される透明性の程度との均衡を取る必要があり、高度な意思決定において解釈可能性の必要性と効率性向上との慎重な比較検討が求められます。

## データプライバシー

個人データを自らコントロールする権利は持続的な原則であり、国際的にも広く受け入れられています。今後数年間で起こると考えられるのは、データプライバシーの維持がより困難になるという点です。PVにおけるAIの発展は、今後さらに急速に進展すると見込まれています。ビッグデータ解析、フェデレーテッドラーニング、ブロックチェーンの導入はデータセキュリティ、相互運用性、および国際的協働を強化することになります。例えば、AI搭載チャットボットやバーチャルアシスタントは、患者および医療従事者 (HCP) とシームレスに対話することで、リアルタイムの副作用 (ADR) 報告を促進すると考えられます。<sup>20,21</sup> 第7章で述べたとおり、個人データの漏えいは頻度を増しており、その一部は極めて大規模なものとなっています。<sup>22</sup> オンラインプラットフォームを用いたコミュニケーションやサービス利用の増加に伴い、収集されたデータの利用方法に関する理解不足や、データ侵害リスクを容認する傾向が一部の国で見られます。侵害は過失から犯罪的意図まで様々な理由で発生しています。医療データの場合、本人の同意なく個人情報が公開されることは、精神的影響、スティグマ化、差別的取扱いのリスクを伴います。

大規模な医療データを収集・連結しようとする圧力は、業務効率化 (臨床的ケアの一貫性確保や医療費管理) および科学的知識の進展の双方の観点から非常に強いものがあり

ます。現時点ではGenAIの利用はまだ初期段階にあります。品質向上と利用拡大が進み生物医学研究、臨床実践、さらには日常生活の多くの領域へ適用が拡大することは確実です。オープン型LLMの利用は、個人データの意図しない開示という特有のリスクを伴い、このリスクが明確になるにつれて今後重要な議論の対象となると考えられます。

社会は、AIによる学習最大化および患者アウトカム向上を目的としたデータの商品化への圧力と、意図しない情報開示からの保護との均衡を取る必要があります。一つの可能性として、データ共有が自動化される一方で、患者または介護者に対する実証可能な価値最大化を義務付けるチェック機構が組み込まれることが考えられます。匿名化を支援するセキュリティ対策として、ブロックチェーン等の技術を活用し、患者識別キーを用いずに完全匿名化を実現しつつ、医療関連データを安全に共有できる仕組みが導入される可能性があります。しかし、適切な規制上の抑制と均衡が存在しなければ、これらのデータが不正利用または悪用される危険性も容易に想定されます。

AIの進化は、安全性データへのアクセスが即時化され、リアルタイムのインサイトを強化し、シームレスなデータ共有を可能にする時代をもたらす可能性があります。これにより安全性評価の迅速性および正確性は大きく向上する可能性があります。一方で、専有権の問題、法的複雑性、または社会的信頼の低下により、データ共有がむしろ制限されるシナリオも同様に現実的です。特に保険業界などにおいてデータが商業的資産としての価値を持つという認識が高まるにつれ、慎重姿勢が強まり、データ流通がさらに制約される可能性があります。さらに、地域間でのデータプライバシー規制およびデータ共有要件の不均一性も、データプライバシーおよび導入における追加的課題となります。

これら相反する要素の均衡を取ることが極めて重要です。信頼を醸成し、データ完全性を確保し、責任あるデータ共有を促進する透明性の高い枠組みを確立することが、公共の信頼を維持しながらAIの潜在能力を最大限に引き出すために不可欠となります。

## 公正性と衡平性

第8章「公正性と衡平性」では、AIシステムの利用によって生じ得る差別的バイアスをどのように特定し、対処し、または予防できるかについて検討しています。

公正性と衡平性とは、患者および医療従事者が、新しく高度なAI技術へ平等にアクセスできることを意味するべきです。

前述のデータプライバシーの項で述べたとおり、AIが広範に利用されるPVにおいては、公平に導入されることを確保することが重要であり、データ共有によって個人が不利益を被らないようにしなければなりません。例えば、より高度なモニタリング、遺伝子プロファイリング、個別化されたリスク評価や介入に伴う医療費の増加、あるいは保険加入や治療における差別につながるリスクを回避する必要があります。

AIは、社会的状況や背景に関わらず、すべての患者に対して安全性データの理解とその関連性を等しく確保し、特定の個人または集団におけるベネフィットとリスクの理解を支援するものであるべきです。

## ガバナンスとアカウントビリティ

第9章では、PVにおけるAIソリューションのライフサイクル各段階に対応したガバナンスフレームワークグリッドを含め、ガバナンスとアカウントビリティについて取り扱っています。

AIの急速な統合は、ほぼリアルタイムでの介入が可能な、動的かつリスクに基づくガバナンスフレームワークの必要性を強く示しています。

これは、AIシステムがより自律的かつ自己判断的になるにつれて、特に重要になります。例えば、自動化された患者または医療従事者（HCP）へのアラート機能では、システム自身が機能および出力を自己監視し、自己検知したアラートに基づいて予防的措置を講じることが想定されます。このような進展は、次のような重要な問いを提起します。すなわち、これら新しい技術能力の進化と並行して、PVにおけるガバナンス、説明責任、およびヒューマン・オーバーサイトはどのように進化していくのかという点です。

理想的には、規制当局およびPV分野の産業リーダーが、AIシステムが責任ある形で開発および導入されることを保証するための頑健な監督メカニズムを確立することが望まれます。データの不正利用を防止し、プライバシー基準を維持し、これらの技術が最終的に患者アウトカムの向上に寄与することを確保するための安全対策が必要です。

PVにおけるAIの自律性が高まるにつれて、適応可能な規制フレームワークの必要性はいっそう強調されます。継続的な監視、能動的な監査、および厳格な査察プロトコルは、リスク低減、患者安全の確保、および公衆衛生の保護のために不可欠となります。これを実現するためには、管理対象となる技術と同様に俊敏かつ応答性の高いガバナンスモデルへの転換が求められます。

## 10.4. 医薬品安全性監視における人工知能の開発と展開に関する今後の検討の結論

AIの普及・導入およびPVへの統合は、本分野にパラダイムシフトをもたらすと考えられ、その中心は迅速またはリアルタイムでのデータ収集、評価および報告に置かれる可能性があります。例えば、症例報告やシグナル検出を目的として、膨大な安全性データを迅速に解析・抽出する能力をもたらします。これは、こうした技術的進歩を活用するための業務の進め方そのものを根本的に変え、プロセスの効率化を促進するとともに、より広い医療環境およびその周辺領域にも変化をもたらす可能性があります。

PVにおけるAIには非常に大きな可能性がある一方で、今後の検討を要する多くの課題も存在します。特に、自律型AIシステムに対する監督のあり方、およびAIがデータプライバシーや倫理的枠組みにどのような影響を及ぼし得るかが重要な論点となります。本報告書で示した指針原則が、引き続き中核的な考慮事項であり続けることは極めて重要です。ただし、それらの原則は、PVおよび医学全般におけるAIの進歩と応用に応じて、進化し適応していく必要があるという理解が前提となります。これは、PVにおけるAI利用が偏りなく、透明で、安全であり続け、不適切な利用や偶発的な危害を防止することを確実にするためです。規制上および倫理上の安全措置を含む適切なヒューマン・オーバーサイトは、適用される技術的進歩と同様に極めて重要です。

## 第10章 – 参考文献

- 1 Ventola CL. The nanomedicine revolution: part 1: emerging concepts. P T. 2012;Sep;37(9):512-525. ([Journal full text](#) accessed 28 April 2025)
- 2 Ventola CL. The nanomedicine revolution: part 2: current and future clinical applications. PT. 2012;Oct;37(10):582-591. ([Journal full text](#) accessed 28 April 2025)
- 3 Ventola CL. The nanomedicine revolution: part 3: regulatory and safety challenges. PT. 2012;Nov;37(11):631-639. <https://doi:10.1007/s13300-012-0003-x> ([Journal full text](#))
- 4 Ventola CL. Big data and pharmacovigilance: data mining for adverse drug events and interactions. PT. 2018; Jun;43(6):340-351. ([Journal full text](#))

- 5 Kjoersvik O, Bate A. Black Swan Events and Intelligent Automation for Routine Safety Surveillance. *Drug Saf.* 2022;May;45(5):419-427. <https://doi.org/10.1007/s40264-022-01169-0>. (Journal full text)
- 6 Shneiderman B. Human-centered artificial intelligence: reliable, safe & trustworthy. [preprint]. arXiv. 2020;Feb23. <https://doi.org/10.48550/arXiv.2002.04087> (Journal full text)
- 7 National Academies of Sciences, Engineering, and Medicine. Opportunities and challenges for digital twins in biomedical research: proceedings of a workshop—in brief. Washington (DC): The National Academies Press; 2023. <https://doi.org/10.17226/26922>. (Journal full text)
- 8 Katsoulakis E, Wang Q, Wu H, et al. Digital twins for health: a scoping review. *npj Digit Med.* 2024;7:77. <https://doi.org/10.1038/s41746-024-01073-0> (Journal full text accessed 19 September 2025)
- 9 Zhang K, Zhou H-Y, Baptista-Hon DT, Gao Y, et al. Concepts and applications of digital twins in healthcare and medicine. *Patterns.* 2024;5(8):101028., <https://doi.org/10.1016/j.patter.2024.101028>. (Journal full text accessed 19 September 2025)
- 10 Chouldechova A, Roth A. A snapshot of the frontiers of fairness in machine learning. *Commun ACM.* 2020;63(5):82–89. <https://doi.org/10.1145/3376898> (Journal full text accessed 28 April 2025)
- 11 Hauben M. Artificial intelligence and data mining for the pharmacovigilance of drug–drug interactions. *Clin Ther.* 2023;45(2):117-133. doi:10.1016/j.clinthera.2023.01.002. (Journal full text accessed 28 April 2025)
- 12 Sardari S, Hheidari A, Ghodousi M, Rahi A, et al. Nanotechnology in tissue engineering: expanding possibilities with nanoparticles. *Nanotechnology.* 2024; 35 392002. DOI 10.1088/1361-6528/ad5cfb. (Journal full text accessed 24 October 2025)
- 13 Zhu X, Wang Z, Teng F. A review of regulated self-organizing approaches for tissue regeneration. *Prog Biophys Mol Biol.* 2021;167:63–78. <https://doi.org/10.1016/j.pbiomolbio.2021.07.006>. (Journal full text accessed 28 April 2025)
- 14 Wang C, He T, Zhou H, Zhang Z, Lee C. Artificial intelligence enhanced sensors - enabling technologies to next-generation healthcare and biomedical platform. *Bioelectron Med.* 2023;Aug2;9(1):17. <https://doi.org/10.1186/s42234-023-00118-1> (Journal full text accessed 28 April 2025)
- 15 Abyzova E, Dogadina E, Rodriguez RD, Petrov I, et al. Beyond Tissue replacement: The Emerging role of smart implants in healthcare. *Mater Today Bio.* 2023 Aug 29;22:100784. doi: 10.1016/j.mtbio.2023.100784. (PubMed accessed 24 October 2025)
- 16 Choudhury A, Asan O. Human factors: bridging artificial intelligence and patient safety. *Proc Int Symp Hum Factors Ergon Health Care.* 2020;Oct5;9(1):211–215. <https://doi.org/10.1177/2327857920091007> (Journal abstract accessed 28 April 2025)
- 17 Bonabeau E, Dorigo M, Theraulaz G. *Swarm intelligence: from natural to artificial systems.* New York: Oxford University Press; 1999; online edn 2020;Nov12. <https://doi.org/10.1093/oso/9780195131581.001.0001> (Abstract accessed 28 April 2025)
- 18 Kurzweil, Ray (2005). *The Singularity is Near: When Humans Transcend Biology.* Viking Press.
- 19 Karakas F. Imagine you are living in the age of singularity. [Internet]. *Creative Adventures.* 2021;75 (Webpage accessed 28 April 2025)
- 20 Chen C, Feng X, Li Y, Lyu L, et al. Integration of large language models and federated learning. *Patterns.* 2024;5(12):101098., <https://doi.org/10.1016/j.patter.2024.101098>. (Journal full text accessed 19 September 2025)
- 21 Shama SN. AI in pharmacovigilance: the dawn of a data-driven safety revolution. [Internet]. *Pharma Focus America;* 2025. (Webpage accessed 19 September 2025)
- 22 Forbes. Cybersecurity stats: facts and figures you should know. [Internet]. New York: Forbes Media; 2024. (Webpage accessed 28 April 2025)

# 付録 1. 用語集

本用語集は、医薬品安全性監視（PV）における人工知能（AI）利用に関連する用語について本書の目的に沿って定義を示すもの、または既存の定義（例：技術的定義）を本書の目的のために簡略化したものです。本用語集に含まれていない追加用語については、CIOMSが作成した「ICH用語および定義集（[Glossary of ICH Terms and Definitions](#)）」ならびにその他関連する用語集を参照してください。

## アカウンタビリティ（Accountability）

アカウンタビリティとは、医薬品安全性監視（PV）の義務を果たすために人工知能システムを導入、運用および管理する組織および／または個人に対して、明確に定義された役割、責任および法的責任が適用されることを指します。これは、関連するステークホルダー（規制当局、ベンダー、ユーザー、開発者、データ提供者、または人工知能システムの方針策定、開発、導入、維持および管理に関与する製薬企業を含むがこれらに限定されない）による適切なガバナンス措置の採用を要求します。これにより、人工知能ライフサイクル全体を通じて運用が想定された範囲内に維持されるとともに、予期しない結果にも対応できることを確保します。

CIOMS Working Group XIV.による提案

## 有害事象（Adverse event）

医薬品が投与された患者または臨床試験参加者において発現した、好ましくないあらゆる医学的事象を指し、当該治療との因果関係を必ずしも有する必要はありません。したがって、有害事象（AE）とは、医薬品（試験薬を含む）の使用と時間的関連を有する、好ましくなく意図されない徴候（異常な検査値を含む）、症状または疾病を指し、当該医薬品（試験薬）との関連の有無を問いません。

出典： Council for International Organizations of Medical Sciences (CIOMS). Glossary of ICH terms and definitions. Geneva: Council for International Organizations of Medical Sciences; 2024. ([Full text accessed 4 April 2025](#))

## 副作用（Adverse reaction）

医薬品に対する有害かつ意図しない反応を指し、当該製品と事象との間に少なくとも合理的な因果関係の可能性が存在することを意味します。

出典改変： Council for International Organizations of Medical Sciences (CIOMS). Glossary of ICH terms and definitions. Geneva: Council for International Organizations of Medical Sciences; 2024. ([Full text accessed 4 April 2025](#))

## AIエージェント (Agent in AI)

AIエージェントとは、環境と相互作用しながらデータを収集し、そのデータを活用してあらかじめ定義された目標を達成するための特定のタスクを実行するソフトウェアプログラムを指します。エージェントは単独で動作することも、共通の目標を達成するために他のエージェントと協働することもあります。

出典改変： Amazon Web Services (AWS). What are AI agents? [Internet]. Seattle (WA): Amazon Web Services; 2024. ([Webpage](#) accessed 15 October 2025)

## 人工知能リテラシー (Artificial intelligence literacy)

人工知能リテラシーとは、AI駆動技術を通じてデジタル社会において理解し、学習し、働くために必要な基本的能力を指します。

出典改変： Ng DTK, Leung JKL, Chu SKW, Qiao MS. Conceptualizing AI literacy: an exploratory review. *Comput Educ Artif Intell.* 2021;2:100041. <https://doi.org/10.1016/j.caeai.2021.100041>. (Journal full text)

## 人工知能システム (Artificial intelligence system)

人工知能 (AI) システムとは、明示的または暗黙的な目的のもとで、受け取った入力から推論を行い、予測、コンテンツ、推奨、または意思決定などの出力を生成し、物理的または仮想的環境に影響を与える機械ベースのシステムを指します。

出典改変： Organisation for Economic Co-operation and Development (OECD). Explanatory memorandum on the updated OECD definition of an AI system. (OECD Artificial Intelligence Papers, No. 8.) Paris: OECD Publishing; 2024. ([PDF](#) accessed 15 October 2025) <https://doi.org/10.1787/623da898-en>.

注記：医薬品安全性監視 (PV) においては、AIシステムおよび関連活動は、医薬品安全性監視、患者安全および規制遵守の強化を目的として使用されます。

## 拡張知能／知能拡張 (Augmented intelligence / Intelligence augmentation)

拡張知能とは、人工知能の補助的役割に焦点を当てた概念であり、人間の知能を置き換えるのではなく、強化（拡張・増幅）するためにAIを活用する考え方を指します。この概念では、人工知能と人間が人間中心のパートナーシップのもとで協働し、それぞれが単独では達成できないタスクを補完的に遂行できることを前提としています。

出典統合：

- Madni AM. Augmented intelligence: a human productivity and performance amplifier in systems engineering and engineered human-machine systems. In: *Systems engineering for the digital age: practitioner perspectives*. Hoboken (NJ): Wiley; 2023; Oct 8. p. 375-391. <https://doi.org/10.1002/9781394203314.ch17> (Chapter abstract)
- World Medical Association (WMA). WMA statement on augmented intelligence in medical care. Ferney-Voltaire (France): World Medical Association; 2019. ([Webpage](#) accessed 3 April 2025)

## 自動化バイアス／自動化コンプレイセンシー（Automation bias or automation complacency）

自動化バイアスおよび自動化コンプレイセンシーは、自動化によって生じる現象の相互に重なり合う形態であり、人間の注意のあり方が中心的役割を果たします。いずれも、自動化された意思決定システムから提示された提案を、人間が非自動化の矛盾する情報よりも優先して信頼・採用してしまう傾向を指し、たとえ後者が正しい場合であっても生じ得ます。これには、自動化出力へ注意が偏る注意バイアス、または特にマルチタスク環境において手動作業が専門家の注意と競合する場合に、自動化出力に対する監視や注意が不十分になる状況が含まれます。

出典統合：

- Parasuraman R, Manzey DH. Complacency and bias in human use of automation: an attentional integration. *Hum Factors*. 2010;Jun;52(3):381-410. <https://doi.org/10.1177/0018720810376055> (Journal full text)
- Cummings ML. Automation bias in intelligent time-critical decision support systems. In: *Decision making in aviation*. Boca Raton (FL): Routledge; 2017;Jul5. p. 289-294. (Chapter abstract accessed 4 April 2025)

## バイアス（Bias）

バイアスとは、測定過程が母集団パラメータの値を過大または過小に推定してしまう傾向を指します。

出典採用： Council for International Organizations of Medical Sciences (CIOMS). Glossary of ICH terms and definitions.

Geneva: Council for International Organizations of Medical Sciences; 2024. (Full text accessed 4 April 2025)

人工知能（AI）においては、バイアスとは、特定の対象、個人、または集団が他と比較して体系的に異なる扱いを受ける状態を指す場合があります（ISO/IEC DIS 22989）。バイアスは研究デザイン、実施、または解析の過程で導入され得ます。代表的な例として選択バイアス（研究対象の選定）、運用バイアス、欠測データを考慮しない解析などが挙げられます。

出典改変： International Medical Device Regulators Forum (IMDRF). Machine learning-enabled medical devices — a subset of artificial intelligence-enabled medical devices: key terms and definitions. Geneva: International Medical Device Regulators Forum; 2021. (Full text accessed 3 April 2025)

さらにAIの利用においては、データやアルゴリズムが既存の社会的不平等を反映または増幅することで、差別的または不公平なAI出力を生じる場合にもバイアスが発生します。

出典改変: University of Saskatchewan. Generative artificial intelligence: glossary of AI-related terms. Saskatoon (SK):

University of Saskatchewan; 2024. (Webpage accessed 4 April 2025)

## ブラックボックスモデル（Black-Box model）

ブラックボックスモデルとは、入力されたデータに基づいて結果を生成するものの、その結果に至る論理や内部処理過程を外部から特定または推定できないAIモデルを指します。

CIOMS Working Group XIV.による提案

## ブラックスワン事象 (Black Swan event)

通常の予測の範囲外にあり、事前には予見不可能であるにもかかわらず、極めて大きな影響をもたらす事象を指します。このような事象は発生後に、人間がその原因を説明可能で予測可能であったかのように解釈してしまう傾向を伴います（すなわち事後的な認識の歪み）。

出典統合：

- Kjoersvik O, Bate A. Black swan events and intelligent automation for routine safety surveillance. *Drug Saf.* 2022;May;45(5):419-427. <https://doi.org/10.1007/s40264-022-01169-0> (Journal full text)
- Taleb NN. Black swans and the domains of statistics. *Am Stat.* 2007;Aug 1;61(3):198-200. [doi:10.1198/000313007X219996](https://doi.org/10.1198/000313007X219996). (Journal full text)

## 事業継続計画 (Business continuity plan)

組織の人員およびインフラ全体、または特に医薬品安全性監視 (PV) の構造および業務プロセスに重大な影響を及ぼし得る事象の予防および復旧のために定められる規定および体制を指します。これには、組織内部、PV業務を共同で実施する組織間、ならびに製薬承認保有者 (MAH) と規制当局との間における緊急時の情報交換が含まれます。

出典改変：Heads of Medicines Agencies (HMA), European Medicines Agency (EMA). Guideline on good pharmacovigilance practices (GVP): Module I – pharmacovigilance systems and their quality systems. London: European Medicines Agency; 2012.v (Full text accessed 3 April 2025)

## 変更管理 (Change management)

変更管理とは、組織構造および／または業務プロセスの変更を計画、実施、および統制するために設計・使用されるプロセス、手法、および技術を指します。方法論は、人、プロセス、および組織文化を中心として構成されます。

一般的に変更管理には、以下の要素が含まれます：リーダーシップ整合、ステークホルダー参画、コミュニケーション、教育訓練、影響評価、継続的改善。

出典改変：International Organization for Standardization (ISO). What is change management: a quick guide. Geneva: International Organization for Standardization; 2023. (Webpage accessed 27 October 2025)

## クラス不均衡 (Class imbalance)

分類タスクにおいてカテゴリー間の件数分布に偏りが存在する状態を指します。この不均衡はモデル性能指標に影響を与えます。例えば、テストケースの99%が同一クラスに属する場合、常に同じ結果を予測するモデルであっても99%の精度を示す可能性があります。

出典採用：European Medicines Agency (EMA). Reflection paper on the use of artificial intelligence (AI) in the medicinal product lifecycle. Amsterdam: European Medicines Agency; 2024. (Full text accessed 3 April 2025)

## クラスタ分析 (Cluster analysis)

データ要素間の類似性に基づいて異なるデータを複数の集合に分割し、単独では認識しにくいパターンを特定するための機械学習手法を指します。

出典派生：Jain AK. Data clustering: 50 years beyond K-means. *Pattern Recognit Lett.* 2010;31(8):651-666. <https://doi.org/10.1016/j.patrec.2009.09.011> (Journal full text)

## コンピュータ化システムバリデーション (Computerized system validation)

コンピュータ化システムの設計段階から廃止（または新システムへの移行）に至るまで定義された要求事項が一貫して満たされていることを確立し、文書化するプロセスを指します。バリデーションのアプローチはリスク評価に基づくべきであり、システムの意図された使用目的、および被験者保護や試験結果の信頼性へ与える潜在的影響を考慮する必要があります。

出典改変：International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). Integrated addendum to ICH E6(R1): guideline for good clinical practice E6(R2). Geneva: International Council for Harmonisation; 2016. ([Full text](#) accessed 3 April 2025)

## 確認バイアス (Confirmation bias)

事前の仮説や想定結果を支持するデータを過度に重視し、それに反する証拠を探索しなかったり、軽視・排除したりする傾向を指します。

出典改変：Elston DM. Confirmation bias in medical decision-making. Journal of the American Academy of Dermatology. 2020;Mar1;82(3):572. <https://doi.org/10.1016/j.jaad.2019.06.1286> ([Journal full text](#))

## 交差検証 (Cross-validation)

機械学習モデルの汎化性能（未知データへの適用能力）を評価し、過学習（overfitting）を防止するために用いられる再サンプリング手法を指します。

出典改変：Berrar D. Cross-validation. Preprint submitted to Encyclopedia of Bioinformatics and Computational Biology, 2<sup>nd</sup> ed. Amsterdam: Elsevier; 2019;542-545. ([Full text](#) accessed 3 April 2025).

注記：開発時に訓練データと検証データを完全に分離する代替手法として用いられ、データをより効率的に活用できます。

## データ匿名化 (Data anonymisation)

直接的および間接的な個人識別子を除去し、さらに技術的保護措置を適用することで、個人の再識別リスクをゼロに近づけることを目的とした個人データ処理プロセスを指します。

出典改変：World Health Organization (WHO). Ethics and governance of artificial intelligence for health: guidance on large multi-modal models. Geneva: World Health Organization; 2024. ([Webpage](#) accessed 3 April 2025)

## データドリフト (Data drift)

本番運用中のモデルが時間経過とともに受け取る入力データの分布が変化する現象を指します。この変化により、モデル性能が低下する可能性があります。データの基礎的特性が変化することで発生し、予測モデルの精度および信頼性に影響を及ぼします。

出典改変：U.S. Food and Drug Administration (FDA). FDA digital health and artificial intelligence glossary – educational resource. Silver Spring (MD): U.S. Food and Drug Administration; 2024. ([Webpage](#) accessed 3 April 2025)

## データプライバシー (Data privacy)

データプライバシーとは、個人情報の保護という個人の基本的権利を守るために講じられる措置を指します。

医薬品安全性監視 (PV) の文脈では、これらの措置は特に機微情報および個人データ (健康情報を含む) の保護を重視します。

CIOMS Working Group XIV.による提案

## 決定木 (Decision tree)

データを複数のサブセットに分類することで、データ間に存在する構造、パターン、および関係性を特定するためのモデルを指します。

出典改変: Dikshit A, Pradhan B, Santosh M. Artificial neural networks in drought prediction in the 21st century: a scientometric analysis. Appl Soft Comput. 2022;114:108080. <https://doi.org/10.1016/j.asoc.2021.108080> (Journal full text)

## ディープラーニング (Deep learning)

複数層から構成されるニューラルネットワーク (人工ニューロン/パーセプトロン [ノード]) を用いる機械学習の一種であり、画像・動画・テキストなどの非構造化データから高次特徴を抽出することを可能にします。

出典: Thirunavukkarasu AJ, Ting DS, Elangovan K, Gutierrez L, Tan TF, Ting DS. Large language models in medicine. Nat Med. 2023;Aug;29(8):1930-1940. <https://doi.org/10.1038/s41591-023-02448-8> (Journal full text)

また、多数の隠れ層 (many hidden layers: 中間層) を持つニューラルネットワークの学習を通じて、階層的で豊かな表現を生成するアプローチを指します。

出典: European Medicines Agency (EMA). Reflection paper on the use of Artificial Intelligence (AI) in the medicinal product lifecycle. 2024. (Full text accessed 3 April 2025)

## 動的 (適応型/継続学習型) AIモデル (Dynamic / adaptive / continual learning AI model)

AIシステムのライフサイクルにおける運用段階において、新たなデータや環境変化への曝露に基づき、継続的に学習または適応を行うAIモデルを指します。

出典改変: International Medical Device Regulators Forum (IMDRF). Machine Learning-enabled Medical Devices - A subset of Artificial Intelligence-enabled Medical Devices: Key Terms and Definitions. 2021. (Full text accessed 3 April 2025)

## 説明可能性 (Explainability)

AIシステムが特定の結果を導いた要因や判断ロジック、またはシステム全体の動作原理について、人間が理解できる程度を指します。

CIOMS Working Group XIV.による提案

## 公正性と衡平性 (Fairness and equity)

公正性 (Fairness) とは、多様な集団または人口における差異を考慮しつつ、比較可能な結果を提供するためにバイアスを回避または軽減することを指します。衡平性 (Equity) とは、集団または人口内の違いを認識し、それらの違いを適切に考慮することで、公正な結果を実現することを意味します。

これには、公正性、平等性、非差別、多様性、正義、および法的適合性という概念への理解と遵守が求められます。 [訳注：3頁訳注参照]

不公正なバイアス、差別的または不当な社会的福祉・環境への影響や結果の回避および軽減は、AIライフサイクル全体を通じて考慮されるべきです。

CIOMS Working Group XIV.による提案

## 偽陰性 (False negative)

本来は対象クラスに属しているにもかかわらず、対象外であると誤って判定されたデータポイントを指します。

CIOMS Working Group XIV.による提案

## 偽陽性 (False positive)

対象クラスに属していないにもかかわらず、対象クラスに属すると誤って判定されたデータポイントを指します。

CIOMS Working Group XIV.による提案

## 特徴量 (Feature)

データの測定可能な属性または特性、あるいはデータ処理や変換によって生成され、モデル学習に使用される要素を指します。

CIOMS Working Group XIV.による提案

## 生成AIアプリケーション (Generative artificial intelligence application)

ユーザーから与えられたプロンプトに基づき、テキスト、画像、動画などの新しいコンテンツを生成したり、分類などの識別タスクを実行したりするために、データセットで学習された人工知能手法を用いるコンピュータ化アプリケーションを指します。

CIOMS Working Group XIV.による提案

## 生成型大規模言語モデル (Generative Large Language Models)

多数のパラメータで学習された確率モデルであり、テキスト生成を目的として設計されたアルゴリズムを用いて自然言語処理を可能にするモデルを指します。

出典改変： Chiarello F, Giordano V, Spada I, Barandoni S, Fantoni G. Future applications of generative large language models: a data-driven case study on ChatGPT. Technovation. 2024;133:103002. <https://doi.org/10.1016/j.technovation.2024.103002> (Journal full text)

## AIガバナンス (Governance for AI)

ガバナンスとは、PVシステムにおけるAIの利用を統制し方向付けるために用いられる人間による管理体制を指します。AIガバナンスフレームワークでは、AIガイディングプリンシプルへの適合を確保するため、リスクマネジメントの実務および方針の実装が求められます。

CIOMS Working Group XIV.による提案

## ハルシネーション (Hallucination)

生成AIにおいて、権威ある情報のように提示されるものの、実際には誤りまたは誤解を招く内容である生成結果を指します。

CIOMS Working Group XIV.による提案

## 人間の主体性 (Human agency)

人間が自らの意思に基づいて選択を行い、その選択を行動へと移す能力を指します。

CIOMS Working Group XIV.による提案

## ヒューマン・イン・コマンド (人間による統括管理) (Human-in-command)

人工知能システムの活動全体を人間が監督する能力を指し、経済的・社会的・法的・倫理的影響を含めて把握したうえで、AIシステムを使用するか、いつ使用するか、どのように使用するかを人間が最終的に判断できる状態を意味します。

出典改変： European Commission. Ethics guidelines for trustworthy AI. Brussels: European Commission; 2019. ([Webpage](#) accessed 3 April 2025)

## ヒューマン・イン・ザ・ループ (Human-in-the-loop, HITL)

人工知能システムのすべての意思決定サイクルにおいて人間が介入可能である能力を指します。

出典採用： European Commission. Ethics guidelines for trustworthy AI. Brussels: European Commission; 2019. ([Webpage](#) accessed 3 April 2025)

## ヒューマン・オン・ザ・ループ (Human-on-the-loop, HOTL)

人工知能システムの設計段階および運用監視において、人間が介入可能である能力を指します。

出典改変： European Commission. Ethics guidelines for trustworthy AI. Brussels: European Commission; 2019. ([Webpage](#) accessed 3 April 2025)

## ヒューマン・オーバーサイト (人間による監視) (Human oversight)

Human oversightとは、PVにおけるAIの設計、導入、運用監視、および分析において期待される人間の役割を指します。AIシステムの性能を管理し、潜在的な問題を検出・軽減するためのフレームワークを必要とします。

CIOMS Working Group XIV.による提案

## 個別症例安全性報告 (Individual Case Safety Report)

報告者により、特定の時点において事象または関心対象となる事例を記述するために提供される完全な情報を指します。本報告には、単一の被験者に関する症例、または複数の被験者群に関する症例の情報が含まれる場合があります。

Council for International Organizations of Medical Sciences (CIOMS). Glossary of ICH terms and definitions. Geneva: Council for International Organizations of Medical Sciences; 2024. ([Full text accessed 4 April 2025](#))

## ナレッジグラフ (Knowledge graph)

オブジェクトの組とそれらを結び付ける関係性から構成されるトリプル（事実）の集合により形成される、異種 (heterogeneous) 情報を統合した知識基盤を指します。これらはグラフ構造およびオントロジー（対象領域におけるすべてのオブジェクト、その属性および関係性を機械可読形式で標準化して表現する意味論的フレームワーク）によってモデル化され、既存データセットを統合することで新たな知見を抽出します。

改変出典： Hauben M, Rafi M. Knowledge graphs in pharmacovigilance: a step-by-step guide. Clin Ther.2024;46(7):538-543. <https://doi.org/10.1016/j.clinthera.2024.03.006> (Journal full text)

## 大規模言語モデル (Large language model)

自然言語における単語間の関係性を学習するために、大規模なテキストデータセットを用いて訓練された深層ニューラルネットワークを利用する人工知能モデルの一種を指します。これには、デコーダを有するモデルおよび有さないモデルの双方が含まれます。

出典： Heads of Medicines Agencies (HMA). European Medicines Agency (EMA). Guiding principles on the use of large language models in regulatory science and for medicines regulatory activities; 2024. ([Full text accessed 4 April 2025](#))

## 機械学習 (Machine learning)

入力データに基づいて出力を生成する数学的構造であるモデルのパラメータを、データから最適化する計算的プロセスを指します。機械学習の手法には、教師あり学習、教師なし学習、強化学習などが含まれ、ニューラルネットワークを用いた深層学習など多様な方法が用いられます。

出典： European Medicines Agency (EMA). Reflection paper on the use of Artificial Intelligence (AI) in the medicinal product lifecycle; 2024. ([Full text accessed 3 April 2025](#))

## AIモデル ((AI) Model)

学習データからパターン（特徴量）を学習し、所定の出力を生成できるよう設計された構造内に、パラメータ（重み）が配置された数学的または計算的手法を指します。

改変出典： European Medicines Agency (EMA). Reflection paper on the use of Artificial Intelligence (AI) in the medicinal product lifecycle; 2024. ([Full text accessed 3 April 2025](#))

## モデルドリフト／概念ドリフト (Model drift (Concept Drift) )

モデルの性能が時間の経過とともに変化し、その結果として性能が向上または低下する現象を指します。

改変出典： Wang S, Schlobach S, Klein M. Concept drift and how to identify it. *J Web Semant.* 2011;9(3):247–265. doi:10.1016/j.websem.2011.05.003 ([Journal full text](#))

## 自然言語処理 (Natural language processing)

コンピュータと人間の言語との相互作用に焦点を当てた人工知能分野を指します。

出典： Thirunavukkarasu AJ, Ting DS, Elangovan K, Gutierrez L, Tan TF, Ting DS. Large language models in medicine. *Nat Med.* 2023;Aug;29(8):1930-1940. <https://doi.org/10.1038/s41591-023-02448-8> ([Journal full text](#))

## ネガティブコントロール (Negative control)

関心対象となるクラスに属さないリアルワールドデータとして抽出されたデータポイント、または人工知能モデルにおいて陽性反応が生じないように意図的に作成されたデータを指します。

CIOMS Working Group XIV.による提案

## ニューラルネットワーク (Neural network)

生物学的神経ネットワークに着想を得た計算システムであり、「パーセプトロン（ノード）」と呼ばれる処理単位から構成され、通常は層構造として配置されます。各ノードは相互に通信しながら入力データに対して変換処理を実行します。

出典： Thirunavukkarasu AJ, Ting DS, Elangovan K, Gutierrez L, Tan TF, Ting DS. Large language models in medicine. *Nat Med.* 2023;Aug;29(8):1930-1940. ([Full text](#) accessed 13 November 2025)

## 非決定論的AI (Non-deterministic AI)

同一の入力であっても常に同一の出力が保証されないAIシステムを指します。これは主として、設計上にランダム性、確率的意思決定、または確率的アルゴリズム (stochastic algorithms) が組み込まれていることによります。

CIOMS Working Group XIV.による提案

## オープン型およびクローズド型大規模言語モデル (Open and Closed Large Language Models)

### クローズドモデル (Closed models)

モデルの重み (weights) が一般公開されず、専有ライセンスのもとでアクセスが制限されているモデルを指します。

### オープンモデル (Open models)

モデルの重みにアクセス可能であり、非専有ライセンスに基づき、改変やさらなる検証・研究が可能なモデルを指します。

改変出典： Xu J, Ding Y, Bu Y. Position: open and closed large language models in healthcare [preprint].arXiv. 2025;Jan17. doi:10.48550/arXiv.2501.09906. ([Journal full text](#))

## 過剰適合 (Overfitting)

訓練データに含まれるノイズや偶然的な特徴まで学習してしまい、その結果、新しいデータに対して適切に一般化できなくなる現象を指します。

改変出典：European Medicines Agency (EMA). Reflection paper on the use of Artificial Intelligence (AI) in the medicinal product lifecycle; 2024. ([Full text](#) accessed 3 April 2025)

## パラメータ／ハイパーパラメータ (Parameter, Hyper-parameter)

機械学習モデル内部の変数であり、通常は学習過程において自動的に更新され、モデル性能を最大化するために調整されます。

ディープラーニングでは、パラメータはニューラルネットワークを構成するノードの「重み (weights)」やデータ変換関数を指します。

出典：Thirunavukkarasu AJ, Ting DS, Elangovan K, Gutierrez L, Tan TF, Ting DS. Large language models in medicine. *Nat Med.* 2023 Aug;29(8):1930-1940. <https://doi.org/10.1038/s41591-023-02448-8> ([Journal full text](#))

ハイパーパラメータは、モデルを構成するために使用されるパラメータを指します。モデルパラメータとは異なり、データ学習から直接推定されるものではなく、機械学習モデルの学習開始前に設定する必要があります。

効果的な機械学習モデル構築のためには、ハイパーパラメータ調整 (hyper-parameter tuning) が重要な工程となります。

改変出典：d from Yang L, Shami A. On hyperparameter optimization of machine learning algorithms: theory and practice. *Neurocomputing.* 2020;Nov20;415:295-316. <https://doi.org/10.1016/j.neucom.2020.07.061> ([Journal full text](#) accessed 15 October 2025)

## 性能劣化 (Performance degradation)

人工知能システムが、過去に達成していた期待または要求性能を維持できなくなり、結果として性能が低下または機能不全となる状態を指します。

CIOMS Working Group XIV.による提案

## 個人データ (Personal data)

「個人データ」とは、識別された、または識別可能な自然人（データ主体）に関するあらゆる情報を指します。氏名、識別番号、位置情報、オンライン識別子、または身体的・生理的・遺伝的・精神的・経済的・文化的・社会的アイデンティティに関係する情報などが含まれます。なお、機微（センシティブ）個人データとは、特別なカテゴリーに分類される個人データを指します。

改変出典：European Parliament, Council of the European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), Art. 4(1). *Off J Eur Union.* 2016;L 119:1-88. ([Webpage](#) accessed 4 April 2025)

## 医薬品安全性監視 (Pharmacovigilance)

医薬品に関連する有害作用またはその他の医薬品関連問題について、検出、評価、理解および予防を行う科学および活動を指します。

出典：Council for International Organizations of Medical Sciences (CIOMS). CIOMS cumulative glossary, with a focus on pharmacovigilance. Version 2.1. Geneva: Council for International Organizations of Medical Sciences; 2024. <https://doi.org/10.56759/occf1297> ([Full text](#))

## 医薬品安全性監視システム (Pharmacovigilance system)

医薬品安全性監視 (PV) に関する法的義務および責任を組織が遂行するために使用されるシステムであり、承認された医薬品の安全性を監視し、そのリスク・ベネフィットバランスの変化を検出することを目的として設計されたものです。

出典： Heads of Medicines Agencies (HMA), European Medicines Agency (EMA). Guideline on good pharmacovigilance practices (GVP) Module I – pharmacovigilance systems and their quality systems. London: European Medicines Agency; 2012. ([Full text](#) accessed 3 April 2025)

## 適合率 (Precision)

参照セットにおいて陽性対照として注釈付けされた取得サンプルの割合であり、正しく分類された陽性対照数を当該クラスに割り当てられた全サンプル数で除した比率として算出されます。適合率は陽性的中率 (positive predictive value : PPV) とも呼ばれます。

出典： Hicks SA, Strümke I, Thambawita V, Hammou M, et al. On evaluation metrics for medical applications of artificial intelligence. *Sci Rep.* 2022;Apr 8;12(1):5979. <https://doi.org/10.1038/s41598-022-09954-8> ([Journal full text](#))

## 陽性対照 (Positive control)

対象クラスに属するとして抽出されたリアルワールドデータの一点、または人工知能 (AI) モデルから陽性応答を意図的に引き出すために作成されたデータです。

CIOMS Working Group XIV.による提案

## 予測モデル (Predictive model)

データを解析してパターンおよび傾向を特定し、入力データに基づいて将来の結果または事象を予測できるようにする機械学習アルゴリズムです。

出典： De Hond AA, Leeuwenberg AM, Hooft L, Kant IM, et al. Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review. *NPJ digital medicine.* 2022;Jan10;5(1):2. <https://doi.org/10.1038/s41746-021-00549-7> ([Journal full text](#))

## 品質マネジメントシステム (Quality management system)

ポリシー、プロセスおよびリソースの枠組みを活用して、製品またはシステムの安全性および有効性を維持し改善する、医薬品安全性監視システムの一部です。

出典： International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). Pharmaceutical quality system Q10. Geneva: ICH; 2008. ([PDF](#) accessed 15 October 2025)

## リアルワールドデータ (Real-world data)

患者の健康状態および/または医療提供に関して、さまざまな情報源から日常的に収集されるデータです。RWDの例には、電子健康記録 (EHR) から得られるデータ、医療請求および請求処理データ、製品および疾患レジストリ由来データ、モバイル機器やウェアラブルを含む患者生成データ、ならびに健康状態の把握に資するその他の情報源から収集されたデータ (例：特定の医療システムで収集された遺伝学的およびその他の生体分子表現型データ) が含まれます。

出典： Council for International Organizations of Medical Sciences (CIOMS). Glossary of ICH terms and definitions. Geneva: Council for International Organizations of Medical Sciences; 2024. ([Full text](#) accessed 4 April 2025)

## 再現率 (Recall)

陽性対照のうち正しく分類された割合であり、正しく分類された陽性対照数を全陽性対照数で除した比率として算出されます。感度 (sensitivity) または真陽性率 (true positive rate : TPR) とも呼ばれます。

出典 : Hicks SA, Strömke I, Thambawita V, Hammou M, Riegler MA, Halvorsen P, Parasa S. On evaluation metrics for medical applications of artificial intelligence. *Sci Rep.* 2022;Apr8;12(1):5979. <https://doi.org/10.1038/s41598-022-09954-8> (Journal full text)

## レッドチーム (Red team)

組織のセキュリティ体制に対して、潜在的な攻撃者による攻撃または悪用能力を模擬することを目的として、正式に権限付与され組織化された集団です。レッドチームの目的は、攻撃が成功した場合の影響を示すとともに、運用環境において防御側 (すなわちブルーチーム) にとって有効に機能する対応を示すことにより、組織のサイバーセキュリティを向上させることです。Cyber Red Team とも呼ばれます。

出典 : National Institute of Standards and Technology (NIST). Glossary [Internet]. Gaithersburg (MD): NIST Computer Security Resource Center; 2025. (Webpage accessed 23 October 2025)

## 再現可能性 (Reproducibility)

同一条件下で解析を繰り返した場合に、一貫した結果を得ることができる能力です。データおよびコンピュータコードを用いて結果を再生成します。

出典 : National Academies of Sciences, Engineering, and Medicine; Policy and Global Affairs; Committee on Science, Engineering, Medicine, and Public Policy; Board on Research Data and Information; Division on Engineering and Physical Sciences; Committee on Applied and Theoretical Statistics; Board on Mathematical Sciences and Analytics; Division on Earth and Life Studies; Nuclear and Radiation Studies Board; Division of Behavioral and Social Sciences and Education; Committee on National Statistics; Board on Behavioral, Cognitive, and Sensory Sciences; Committee on Reproducibility and Replicability in Science. Reproducibility and replicability in science. Washington (DC): National Academies Press (US); 2019;May 7. Chapter 3, Understanding reproducibility and replicability. (Chapter full text accessed 27 October 2025)

## リスクに基づくアプローチ (Risk-based approach)

リスクに基づくアプローチは、人工知能 (AI) システムがもたらし得る潜在的な危険性を認識し、コアとなる医薬品安全性監視 (PV) 業務の実行において、異なるユースケースが異なる種類および水準のリスクを伴うことを認識する考え方です。そのため、プロセス管理を考慮しながら、医薬品安全性監視システムの挙動および結果に悪影響を及ぼす可能性のあるリスクを特定し、優先順位付けし、管理するリスク評価が必要となります。リスクは、想定される影響の大きさと負の結果が生じる可能性の双方によって特徴付けられます。

このアプローチはまた、リスクの大きさに見合った方法で誤りおよびバイアスを特定し低減する手順を支援します。これはAIシステムの実装戦略にも影響を与え、一般的に特定されたリスクに見合う形で実施されるべきものとされます。

CIOMS Working Group XIV.による提案

## 頑健性 (Robustness)

データの変動を考慮しながら、システムが意図した目的を信頼性をもって達成できる能力です。

CIOMS Working Group XIV.による提案

## データの二次利用 (Secondary Use of Data)

当初収集された目的とは異なる目的のために既存データを使用することを指します。AIにおいては、モデルの学習または検証を目的として使用されるデータがこれに含まれる場合があります。

出典： Council for International Organizations of Medical Sciences (CIOMS). Glossary of ICH terms and definitions. Geneva: Council for International Organizations of Medical Sciences; 2024. ([Full text accessed 4 April 2025](#))

## セマンティックベクトル (Semantic vector)

単語、語句、または文書を識別子として数学的に表現したものであり、その識別子が高次元空間内で占める位置が当該単語または語句の意味や関係性を表します。これにより人工知能 (AI) システムは意味に基づいて異なるテキスト間の関連性や類似性を理解できるようになります。

出典： Cohen T, Widdows D. Empirical distributional semantics: methods and biomedical applications. J Biomed Inform.2009;Apr1;42(2):390-405. <https://doi.org/10.1016/j.jbi.2009.02.002> ([Journal full text](#))

## 感度分析 (Sensitivity analysis)

入力データまたはモデルパラメータの変化が人工知能 (AI) モデルの出力にどのような影響を与えるかを評価するために用いられる評価手法です。

CIOMS Working Group XIV.による提案

## シグナル (Signal)

一つまたは複数の情報源 (観察および実験を含む) から生じる情報であり、介入と事象または関連する事象群 (有害または有益のいずれも含む) との間に、新たな潜在的因果関係、または既知の関連性に関する新たな側面が存在することを示唆するものです。さらに、その可能性が十分に高く、検証のための追加対応を正当化すると判断される場合を指します。

出典： Council for International Organizations of Medical Sciences (CIOMS). Practical aspects of signal detection in pharmacovigilance. Geneva: Council for International Organizations of Medical Sciences; 2010. ([Full text accessed 15 October 2025](#))

## 静的AIモデル (Static AI model)

導入後に変更されないAIモデルです。

CIOMS Working Group XIV.による提案

## 教師あり学習 (Supervised learning)

学習時にラベル付けされたデータを利用する機械学習を指します。(ISO/IEC DIS 22989)

出典：International Medical Device Regulators Forum (IMDRF). Machine learning-enabled medical devices — a subset of artificial intelligence-enabled medical devices: key terms and definitions. Geneva: International Medical Device Regulators Forum; 2021. ([Full text](#) accessed 3 April 2025)

## テストデータセット (Test dataset)

機械学習モデルの学習時には一切提示されず、モデルが何を学習したかを検証するために使用されるデータの部分集合です。(ISO/IEC DIS 22989 を改変)

出典：International Medical Device Regulators Forum (IMDRF). Machine learning-enabled medical devices — a subset of artificial intelligence-enabled medical devices: key terms and definitions. Geneva: International Medical Device Regulators Forum; 2021. ([Full text](#) accessed 3 April 2025)

## トレーサビリティ (AI) (Traceability (AI))

人工知能 (AI) モデルおよびその生成結果を作成するために使用されたデータ、プロセス、および分類を追跡し記録できる能力です。

CIOMS Working Group XIV.による提案

## トレーニング (Training)

学習データを使用して、機械学習アルゴリズムに基づく機械学習モデルのパラメータを確立または改善することを目的としたプロセスです。(ISO/IEC DIS 22989 を改変)

出典：International Medical Device Regulators Forum (IMDRF). Machine learning-enabled medical devices — a subset of artificial intelligence-enabled medical devices: key terms and definitions. Geneva: International Medical Device Regulators Forum; 2021. ([Full text](#) accessed 3 April 2025)

## トレーニングデータセット (Training dataset)

機械学習において特に使用されるデータであり、機械学習アルゴリズムが与えられた課題に対応するモデルを抽出するための基礎データとして機能します。

出典：European Medicines Agency (EMA). Reflection paper on the use of Artificial Intelligence (AI) in the medicinal product lifecycle; 2024. ([Full text](#) accessed 3 April 2025)

## 透明性 (Transparency)

AIに関する透明性とは、組織間または個人間で情報を開示することを指します。これには、トレーサビリティを可能にするためにAIシステムのライフサイクル（すなわち設計開発、評価、導入、運用、再学習、保守および廃止）に関する関連文書を共有することおよびステークホルダーがAIシステム、その利用方法、リスク、限界、想定される便益ならびに自身の権利への影響について一般的理解を得るのに十分な情報を提供することが含まれます。

CIOMS Working Group XIV.による提案

## 教師なし学習 (Unsupervised learning)

学習時にラベル付けされていないデータを利用する機械学習を指します。(ISO/IEC DIS 22989)

出典：International Medical Device Regulators Forum (IMDRF). Machine learning-enabled medical devices — a subset of artificial intelligence-enabled medical devices: key terms and definitions. Geneva: International Medical Device Regulators Forum; 2021. ([Full text](#) accessed 3 April 2025)

## 妥当性 (Validity)

妥当性とは、システムが許容可能な範囲内で意図した目的を達成することを意味します。そのためには、許容される性能水準を事前に定義し、モデル学習および／または試験に適切なデータを選択し、現実的な環境においてモデル性能を評価し、さらにシステムを継続的な品質保証プロセスに組み込むことが求められます。

CIOMS Working Group XIV.による提案

## バリデーションデータセット (Validation dataset)

ハイパーパラメータの調整、またはアルゴリズム上の選択（ルール設計など）を検証するために使用されるデータです。

出典：International Organization for Standardization (ISO). ISO/IEC DIS 22989. Information technology— artificial intelligence - artificial intelligence concepts and terminology. Geneva: International Organization for Standardization; 2022. ([Webpage](#) accessed 4 April 2025)

## ゼロショット学習 (Zero-shot learning)

当該タスクの過去の例に一切触れることなく、タスクを実行できるように開発された人工知能 (AI) を指します。

出典：Thirunavukkarasu AJ, Ting DS, Elangovan K, Gutierrez L, Tan TF, Ting DS. Large language models in medicine. Nat med. 2023;Aug;29(8):1930-1940. <https://doi.org/10.1038/s41591-023-02448-8> ([Journal full text](#))

# 付録 2. 指針原則の比較表

表10: CIOMS WORKING GROUP XIVの地域および国の政府機関、国際機関における人工知能の指針となる原則の比較 – 原則の抜粋

出典：CIOMS Working Group XIV

原則	EU <sup>12</sup>	オーストラリア <sup>3</sup>	カナダ <sup>4</sup>	シンガポール <sup>5</sup>	英国 <sup>6</sup>	米 <sup>7</sup>	PAHO <sup>8</sup>	WHO <sup>9</sup>	OECD <sup>10</sup>	
地域および各国政府機関ならびに国際機関における原則の例	AIシステムは、人間の尊厳と自律性という原則に基づき、人間の主体性および人間の意思決定を支援するものであるべきです。	AIシステムが個人、コミュニティ、または環境に重大な影響を及ぼす場合には、人々がAIシステムの利用またはその結果に異議を申し立てることができるという適時の手続が存在すべきです。	ヒューマン・オーバーサイトとは、高影響AIシステムが運用を管理する人々が実質的な監督を行うことができるように設計および開発されるべきことを意味します。これは、状況に応じた適切な解釈可能性の水準が含まれます。	AIシステムは、AIライフサイクル全体を通じて、頑健かつ安全で保護されべきであり、リスクは継続的に特定、評価および管理されるべきです。適切な場合には、利用者は、影響を受ける第三者およびAIライフサイクルに関与する主体は、有害な結果をもたらす、または問題を生じさせるAIの判断または結果に対して異議を申し立てることができるべきです。	自動化システムは、「展開前テスト、継続的なモニタリング、および明確な組織的監督」を備えるよう設計・開発されるべきです。	自動化された意思決定に対する人々のおよびレビュエーションの正式なプロセスは必須です。	自律性の原則は、機械の自律性の拡張が人間の自律性を損なうことにならないことを要求します。医療においては、人間が医療システムおよび意思決定を完全に管理し続けるべきであることを意味します。ヒューマン・オーバーサイトはAIシステムに関連するリスクに依存する場合がありますが、常に実質的なものでなければならず、そのため人間の価値観および倫理的配慮を効果的かつ透明に監視することを含むべき	自動化された意思決定に対する人々のおよびレビュエーションの正式なプロセスは必須です。	自律性の原則は、機械の自律性の拡張が人間の自律性を損なうことにならないことを要求します。医療においては、人間が医療システムおよび意思決定を完全に管理し続けるべきであることを意味します。ヒューマン・オーバーサイトはAIシステムに関連するリスクに依存する場合がありますが、常に実質的なものでなければならず、そのため人間の価値観および倫理的配慮を効果的かつ透明に監視することを含むべき	AIシステムが過度の害を引き起こすおそれがある場合、または望ましくない挙動を示す場合には、必要に応じて安全に上書き、修復および/または廃止できるように、適切な仕組みが整備されています。

原則	地域および各国政府機関ならびに国際機関における原則の例								
	EU <sup>1,2</sup>	オーストラリア <sup>3</sup>	カナダ <sup>4</sup>	シンガポール <sup>5</sup>	英国 <sup>6</sup>	米国 <sup>7</sup>	PAHO <sup>8</sup>	WHO <sup>9</sup>	OECD <sup>10</sup>
<b>妥当性と 頑健性</b>	技術的頑健性とは、AIシステムがリスクに対する予防的アプローチに基づいて開発された意図した信頼性をもった動作し、意図した意図しないまたは予期しない害を最小化するるとともに、可能な限りそれを防止することとを要求します。さらに、運用環境の変化や、AIシステムと敵対的に相互作用し得る他の主体（人間または人工的）が存在する場合にも適用されるべきです。	AIシステムは、その意図された目的に従って信頼性をもって動作すべきです。	妥当性とは、高影響AIシステムが意図された目的に沿って一貫して性能を発揮することとを意味します。	シンガポールは、AIに関する主体がシステムの機能、耐性およびセキュリティに定期的な試験またはデューデリジェンスをどのように実施できるかを検討すべきです。	AI介入は、信頼性、再現可能性、公正性、誠実性および説明責任を含む科学的な最良実務に従うべきです。	AIに関する主体がシステムの機能、耐性およびセキュリティに定期的な試験またはデューデリジェンスをどのように実施できるかを検討すべきです。	AI介入は、信頼性、再現可能性、公正性、誠実性および説明責任を含む科学的な最良実務に従うべきです。	すべてのAIシステムは、安全性および有効性の基準を満たしていることを確保するため、当該技術が使用される環境において厳格に試験されるべきです。検証および妥当性確認には、AI技術の前提条件、運用プロトコル、データ特性および出力判断が含まれるべきです。	AIシステムは、そのライフサイクルを通じて健全かつ安全に機能しなければならず、潜在的なリスクは継続的に評価および管理されるべきです。

原則	EU <sup>1,2</sup>	オーストラリア <sup>3</sup>	カナダ <sup>4</sup>	シンガポール <sup>5</sup>	英国 <sup>6</sup>	米国 <sup>7</sup>	PAHO <sup>8</sup>	WHO <sup>9</sup>	OECD <sup>10</sup>
データプライバシー	<p>危害防止の原則にはプライバシーが含まれます。プライバシーは基本的な人権であり、とりわけAIシステムの影響を受けやすい権利です。プライバシーへの危害防止には、使用されるデータの品質および完全性、AIシステムが導入される領域に照らした関連性、アクセス手順、ならびにプライバシーを保護する方法でデータを処理する能力を含む適切なデータガバナンスが必要です。</p>	<p>AIシステムは、プライバシー権およびデータ保護を尊重し維持するとともに、データの安全性を確保すべきです。</p>			<p>AI開発者および導入主体は、それぞれの責任範囲において、AIライフサイクル全体を通じてサイバーセキュリティリスクを低減し、これに対する耐性を構築することが推奨されます。AI開発者および導入主体は、AI製品およびサービスの悪意ある利用または犯罪性の利用について検討し、可能な限りこれを低減することが推奨されます。</p>	<p>「データプライバシーに関する保護は既定として組み込まれ、データ収集が合理的な期待に適合していること、ならびに特定の状況に厳密に必要なデータのみが収集されることを確保することを含みます。」</p>	<p>データ利用におけるプライバシーおよび安全性は、すべてのAI開発の基盤でなければなりません。</p>	<p>データ保護法は「権利基盤アプローチ」に基づいており、個人の権利を保護すると同時に、データ管理者に対する義務を確立する形でデータ処理を規制する基準を提供します。</p>	

地域および各国政府機関ならびに国際機関における原則の例		米国 <sup>7</sup>	PAHO <sup>8</sup>	WHO <sup>9</sup>	OECD <sup>10</sup>	
原則	透明性	シンガポール <sup>5</sup>	英国 <sup>6</sup>	カナダ <sup>4</sup>	オーストラリア <sup>3</sup>	EU <sup>1,2</sup>
		AI開発者および導入主体は、それぞれの責任範囲において、適切な透明性および説明可能性の措置を実施することが推奨されます。	AI開発者および導入主体は、それぞれの責任範囲において、適切な透明性および説明可能性の措置を実施することが推奨されます。	透明性とは、高影響AIシステムがどのように使用されているかについて、適切な情報を公衆に提供することです。提供される情報は、システムの能力、限界および潜在的な影響が理解できるのに十分なものであるべきです。	人々がAIによって大きな影響を受ける場合、またAIシステムが自身と相互作用していること、透明性および責任ある情報開示が確保されるべきです。	透明性は、1) トレーサビリティ、2) 説明可能性、3) AI システムの限界に関するコミュニケーションといった三つの要素を包含します。
		AIアルゴリズムの開発においては、常に透明性のあるアプローチが採用され、それが適切に伝達されなければなりません。可能な限りすべてを公開および共有可能とするべきであり、オープン性の概念およびそれを支えるツールは、あらゆるAI開発における重要な成功要因でなければなりません。	AIは、開発者、利用者および関係者にとって理解可能であるべきです。理解可能性を確保するための二つの主要なアプローチは、AI技術の透明性および説明可能性を向上させることです。	この原則は、AIシステムに關する透明性および責任ある情報開示を促進し、人々がAIと関与していることを理解し、その結果に異議を申し立てることを可能とすることを目指しています。		

原則	地域および各国政府機関ならびに国際機関における原則の例	OECD <sup>10</sup>	WHO <sup>9</sup>	PAHO <sup>8</sup>	米国 <sup>7</sup>	英国 <sup>6</sup>	シンガポール <sup>5</sup>	カナダ <sup>4</sup>	オーストラリア <sup>3</sup>	EU <sup>1,2</sup>	アカウンタビリティ
		AIシステムを開発、導入または運用する組織および個人は、適切な機能について説明責任を負うべきです。	AI技術は特定の作業を実行しますが、それらの作業を実行可能であり、適切な条件下で使用されることを確保する責任は人間の関係者にあります。機関は法的責任を負うだけでなく、アルゴリズムがどのようにに結果を生成するかを詳細に説明することが困難な場合であっても、使用するアルゴリズムによって下された判断について責任を引き受ける義務があります。	妥当性と頑健性を参照。	「自動化システムが機能しない場合、誤りを生じた場合、またはその影響について異議申立てや争いを希望する場合には、代替手段およびエスカレーション手続によって、適時に人間による検討および救済を受けられるようになります。」	ガバナンスを参照。	AI医療機器（AI-MD）に使用される適切な設計について、開発者が責任を負うべきであり、医療提供のためにAI-MDを使用する組織は、AI-MDを導入する判断およびその使用により生じる臨床結果について責任を負います。これは安全な医療提供を確保する責任を含みます。他の医療機器の導入と同様に、AI-MDの使用によって、導入機関または個々の医療専門職が適切かつ安全な医療提供に果たして負う責任が変わるものではありません。	説明責任とは、高影響AIシステムが使用される状況において、すべての法的義務への適合を確保するため、システム体制に必要なガバナンス体制を組織が整備しなければならないことを意味します。これには、実施された方針、プロセスおよび措置の積極的な文書化が含まれます。	AIシステムのライフサイクルの各段階に責任を有する者は特定可能でなければならず、AIシステムの結果について説明責任を負うべきであり、またAIシステムに対するユーザーサイトが可能でなければなりません。	説明責任の原則は、AIシステムの開発、導入および/または使用に関する責任を確保するために、組織の仕組みを整備することを求めます。	

原則	地域および各国政府機関ならびに国際機関における原則の例								
	EU <sup>1,2</sup>	オーストラリア <sup>3</sup>	カナダ <sup>4</sup>	シンガポール <sup>5</sup>	英国 <sup>6</sup>	米国 <sup>7</sup>	PAHO <sup>8</sup>	WHO <sup>9</sup>	OECD <sup>10</sup>
社会的ウェルビーイング	AIシステムの利用がサイクル全体を通じて、より広い社会、他の感覚を有する存在および環境もスーターとして考慮されるべきです。教育、労働、ケアまたは娯楽など生活のあらゆる領域において社会的AIシステムへの広範な接触は、社会的主体性の理解を変化させたり、社会的関係や結びつきに負の影響を及ぼしたりする可能性があります。	AIシステムは、個人、社会および環境に利益をもたらすべきです。		AI医療機器 (AIMD) の設計、開発および実装においては、患者の利益 (安全性および福祉を含む) が保護されることと安全対策が整備されるべきです。			行動および解決策は人間中心でなければならず、AIのみで使用されるべきではありません。公衆衛生を支援する多くの技術の一つとして、AIは個人の権利を尊重すべきです。	AI技術は人々に害を与えません。導入前に安全性、正確性および有効性に関する規制要件を満たす必要があります。品質管理および品質改善を確保するための措置が整備されるべきです。そのため、資金提供者、開発者および利用者には、AI技術が設計どおりに機能していることを確認し、個々の患者または集団に対して有害な影響がないかを評価するため、AIアルゴリズムの性能を継続的に測定および監視する責任を負います。	本原則は、信頼できるAIが、個人、社会および地球の主体としての成長および繁栄に貢献し、世界的な開発目標の推進に寄与する可能性を強調するものです。

地域および各国政府機関ならびに国際機関における原則の例									
原則	EU <sup>1,2</sup>	オーストラリア <sup>3</sup>	カナダ <sup>4</sup>	シンガポール <sup>5</sup>	英国 <sup>6</sup>	米国 <sup>7</sup>	PAHO <sup>8</sup>	WHO <sup>9</sup>	OECD <sup>10</sup>
環境的ウェルビーイング	社会的ウェルビーイングを参照。	社会的ウェルビーイングを参照。						AIシステムは、その生態学的フットプリントを最小化し、エネルギー効率を高めるように設計されるべきであり、それにより、AIの利用が、地球の環境、気候変動に対する人間の影響を軽減しようとする社会の取り組みと整合するものとなるべきです。	社会的ウェルビーイングを参照。

原則	地域および各国政府機関ならびに国際機関における原則の例										
	EU <sup>1,2</sup>	オーストラリア <sup>3</sup>	カナダ <sup>4</sup>	シンガポール <sup>5</sup>	英国 <sup>6</sup>	米国 <sup>7</sup>	PAHO <sup>8</sup>	WHO <sup>9</sup>	OECD <sup>10</sup>		
公正性と 公平性	AIシステムのライフサイクル全体を通じて、包摂性および多様性を確保できるよう努めます。 AIシステム（学習および運用の双方）は、意図しない歴史的不完全なデータ、および不適切なバイアスの影響を受けやすくなります。	AIシステムは包摂的かつ利用可能であり、コミュニティまたは差別を伴って生じさせません。	公正性と公平性は、差別的結果が生じ得る可能性を認識し、高影響AIシステムを構築することを通じて、個人および集団に対する差別的結果を低減するため、適切な措置が講じられなければなりません。	AI医療機器（AI-MD）の実装は、人種や性別など異なる人口統計学的特性に基づき、患者に対して公正な臨床的影響を生じさせません。	AIシステムは、個人または組織の法的権利を損なう不当に差別を行ったり、不正な市場結果を生じさせたりしてはなりません。AIライフサイクルのすべての段階に関する主体は、システムの用途、結果および適用される法令に照らし、適切な公正性の定義を考慮すべきです。	「アルゴリズムからの保護には、システムにおける積極的な公平性評価、代表性のあるデータの使用、人口統計学的特性の代理変数からの保護、ならびに障害を有する人々へのアクセシビリティ確保が含まれます。」	「公衆衛生におけるあらゆるAIの取り組みにおいては、影響の双方における公正性、平等性および包摂性が常に基礎とならなければなりません。議論、開発および実装は、人間の尊厳、善行、無危害および正義という国際的に合意された倫理原則に基づいて行われなければなりません。」	包摂性とは、医療において使用されるAIが、年齢、性別、所得、能力またはその他の特性にかかわらず、可能な限り広く適切かつ公平な利用およびアクセシビリティを促進することです。 AI開発者は、設計、実装および使用において生じ得るバイアスと、それが個人および社会にもたらす可能性のある害について認識しておくべきです。	AIシステムは、法、人権、民権的価値および多様性を尊重する形で設計されており、公正で正義に基づく社会を確保するための適切な保護措置を含みます。		

原則	EU <sup>1,2</sup>	オーストラリア <sup>3</sup>	カナダ <sup>4</sup>	シンガポール <sup>5</sup>	英国 <sup>6</sup>	米国 <sup>7</sup>	PAHO <sup>8</sup>	WHO <sup>9</sup>	OECD <sup>10</sup>
説明可能性	説明可能性とは、AIシステムの技術的プロセスおよびAIシステムが行う判断または予測の根拠の双方を説明できる能力を指します。	透明性を参照。		AI医療機器（AI-MD）による判断または推奨は、説明可能かつ再現可能であるよう努めべきです。説明可能性の水準は、最終利用者の期待およびAI-MDに伴うリスクの程度によって異なります。説明可能性が利用者の期待を満たすことを確保するため、AI-MDの開発または導入の過程において最終利用者との協議すべきです。	透明性を参照。	「自動化システムは、技術的に妥当であり、意味が明確で有用な説明を提供すべきであり、システムを理解する必要がある利用者、運他の関係者に対して理解可能でなければなりません。また、その説明は状況に基づきリスクレベルに応じて調整され、平易な言葉で提示され、通知および説明の明確性と品質の評価は、可能な限り公開されるべきです。」		透明性を参照。	透明性を参照。

地域および各国政府機関ならびに国際機関における原則の例

地域および各国政府機関ならびに国際機関における原則の例		シンガポール <sup>5</sup>	英国 <sup>6</sup>	米国 <sup>7</sup>	PAHO <sup>8</sup>	WHO <sup>9</sup>	OECD <sup>10</sup>		
<b>原則</b>									
<b>安全性</b>	EU <sup>1,2</sup> 妥当性と頑健性を参照。	オーストラリア <sup>3</sup> 妥当性と頑健性を参照。	カナダ <sup>4</sup> 安全性とは、高影響AIシステムについて、合理的に予見可能な誤使用を含め、システム使用によって生じ得る危害を特定するため、事前に評価を実施しなければならぬことを意味します。危害のリスクを低減するための措置が講じられなければならない。	シンガポール <sup>5</sup>	英国 <sup>6</sup> AI導入主体（その責任範囲において）および最終利用者およびAI製品およびサービスの安全性について十分な情報に基づいた判断を行えるようにすべきです。適切にリスクを特定、監視伝達および対応することにより、それぞれの責任範囲における安全性関連リスクの水準を伝達すべきです。	米国 <sup>7</sup> 「自動化システムは、意図しないが予見可能な使用または影響から生じる危害から利用者を積極的に保護するよう設計されるべきです。」	PAHO <sup>8</sup>	WHO <sup>9</sup> 危害防止には、AI技術の使用が精神的または身体的な危害を引き起こさないことが求められます。	OECD <sup>10</sup> 妥当性と頑健性を参照。
<b>ガバナンス</b>	データプライバシーポリシーを参照。					人権基準、データ保護法および倫理原則は、開発者、政府、医療提供者および患者分野におけるAI利用を指導、規制および管理するために不可欠です。			

## 付録 2 – 参考文献

- 1 European Medicines Agency (EMA), Committee for Medicinal Products for Human Use (CHMP). Reflection paper on the use of artificial intelligence (AI) in the medicinal product lifecycle. Amsterdam: European Medicines Agency; 2024; Sep 9 ([PDF](#) accessed 27 April 2025)
- 2 European Commission, Directorate-General for Communications Networks, Content and Technology. The assessment list for trustworthy artificial intelligence (ALTAI) for self-assessment. Luxembourg: Publications Office of the European Union; 2020., <https://data.europa.eu/doi/10.2759/002360> ([PDF](#) accessed 27 April 2025)
- 3 Australian Government, Department of Industry, Science and Resources. Australia's artificial intelligence ethics principles. Canberra: Australian Government; [year unknown] ([Webpage](#) accessed 27 April 2025)
- 4 Government of Canada. The Artificial Intelligence and Data Act (AIDA) – companion document. Ottawa: Government of Canada; 2022. ([Webpage](#) accessed 27 April 2025)
- 5 Ministry of Health, Singapore. Artificial intelligence in healthcare guidelines. Singapore: Ministry of Health; 2021. ([PDF](#) accessed 27 April 2025)
- 6 Department for Science, Innovation and Technology (DSIT). Implementing the UK's AI regulatory principles: initial guidance for regulators. London: UK Government; 2024 Feb. ([PDF](#) accessed 27 April 2025)
- 7 White House Office of Science and Technology Policy (OSTP). About this document: Blueprint for an AI Bill of Rights [Internet]. © 2025; ([Webpage](#) accessed 15 October 2025)
- 8 Pan American Health Organization (PAHO). Artificial intelligence in public health: digital transformation toolkit. Washington (DC): Pan American Health Organization; 2021. (PAHO/EIH/IS/21-011). ([PDF](#) accessed 27 April 2025)
- 9 World Health Organization (WHO). Ethics and governance of artificial intelligence for health: WHO guidance. Geneva: World Health Organization; 2021. Licence: CC BY-NC-SA 3.0 IGO ([Website](#) accessed 27 April 2025)
- 10 Organisation for Economic Co-operation and Development (OECD). OECD AI principles overview. © 2025. ([Website](#) accessed 27 April 2025)

## 付録 3. ユースケース

---

本付録に示すユースケースは、医薬品安全性監視（PV）における人工知能（AI）活用の実践的事例として提示されるものです。各事例は、PVの多様なサブドメインにおけるAI応用の可能性に関するインサイトを提供しAI手法の統合に際しての方法論、限界、および関連する性能を示しています。各ユースケースは、現在のPVライフサイクル全体における特定のAIソリューションを例示しており、AIがいかにかワークフローを変革し、効率性を高め、究極的に患者の安全性向上に貢献できるかについて、読者に実践的な例を示すものです。

これらのユースケースを適切に活用するためには、本文書に記載された諸原則およびそれらの一部がユースケースの開発・実装においてどのように指針を与えてきたかを理解することが重要です。なお、本ユースケースは本文書に先行して作成されたものであり、すべての原則が考慮されているとは限らないことに留意することが重要です。

読者においては、各ケーススタディの背景、目的、および成果を分析し、自らの組織のニーズに応じた有意義なインサイトを得ることが推奨されます。本編に概説されているガバナンス・フレームワークと各AIソリューションとの整合性を体系的に評価することにより、ステークホルダーはAI統合におけるベストプラクティスや潜在的な落とし穴を特定することができ、PV分野におけるAI技術の採用に向けた責任あるアプローチを促進することが可能となります。

さらに、これらのユースケースは、中核となる主要な指針原則を遵守することの重要性を明らかにしています。これらの事例を検討することにより、組織は自らのPV（医薬品安全性監視）システムの完全性を保護しつつ、イノベーションを推進するための貴重な視点を得ることができます。

# ユースケース A: ケース処理のための大規模言語モデルデータ抽出

出典：<sup>1</sup>

PV領域：ICSR Processing

## A1. ビジネス上の背景および課題

製造販売後安全管理の基準（GVP）は、医薬品のライフサイクル全体を通じて安全性および有効性を確保するために、欧州医薬品庁（EMA）によって確立されたガイドラインおよび基準の体系を指します。これらの実務は、有害反応の体系的な監視、評価および管理に不可欠です。製薬企業は公衆衛生を保護するため、医薬品に関連する潜在的な有害反応の報告に対応し、潜在的リスクを迅速に特定し対処しなければなりません。近年、症例報告数が大幅に増加していることにより、ケース受付および処理業務は、件数の増加だけでなく、非構造化テキストやスキャン文書を含む多様なデータソースへの対応や、限られた人員での突発的な大量流入への対応など、複雑な課題に直面しています。対象となるデータポイントは、単純な人口統計情報から複雑な検査値に至るまで多岐にわたります。このため、報告された有害事象名、報告者資格、国名および特定用語などを、事前定義された注釈に基づいて識別および分類する固有表現認識（Named Entity Recognition）のような比較的単純な技術的手法では実環境においてケース受付および処理業務を一貫して改善することはできませんでした。ケース受付および処理における大規模言語モデル（LLM）の活用は、品質を損なうことなくプロセスを高度化できる可能性を有しています。ただし、LLMベースのツールについては、モデルドリフト（model drift）を効果的に監視するため、定期的な再評価を実施する必要があります。さらに、学習およびキャリブレーションの過程においては、識別可能なデータの取扱いがGDPRおよびHIPAA規制に適合することを確保し、データプライバシーの保護およびコンプライアンス維持を図らなければなりません。

## A2. ソリューション

ある製薬企業は、ケース受付業務におけるLLM活用の実現可能性および定量的・定性的なビジネス影響を評価するため、概念実証（PoC）研究を実施しました。具体的には、規制およびコンプライアンス要件を考慮しながら、ケース受付および処理における原資料からのデータ抽出にLLMを適用しました。

選定された原資料を処理し、事前定義された情報項目を抽出するために、三段階からなる半自動処理パイプラインが構築されました。このパイプラインは、（1）LLM（OpenAIのGPT-4）への入力を統一する前処理工程、（2）情報構造化をLLMに指示するとともに原資料内の情報位置に関する手掛かりを与えるJSON形式の抽出テンプレート、（3）モデル出力を事前定義値が存在するフィールドへ適合させる後処理工程から構成されました。原資料の匿名化コピーには、抽出された重要用語の参照およびハイライトが付与されました。

ケース受付にLLMを使用した場合のビジネス影響を評価するため、代表的な症例セットが選定されました。また、（a）完全手作業による処理と、（b）LLM抽出パイプラインの結果により事前入力されたフィールドを用いて補助された手作業処理の性能を比較する目的で、グラフィカルユーザーインターフェース（GUI）が設計されました。

<sup>1</sup> JSON = JavaScript Object Notation

4名の経験豊富な専門担当者が無作為に (a) または (b) の処理方式へ割り当てられました。各原資料について処理時間が記録され、代表的フィールド集合の抽出に要する総処理時間が算出されました。

### A3. 結果

本研究では、ケース受付および処理業務における大規模言語モデル (LLM) の導入から二つの主要な結果が得られました。

第一の結果は、LLMモデルの性能に関するものであり、本研究の対象となったすべての原資料について、抽出された全フィールドの一致スコアを測定し、症例カテゴリーごとに平均化して評価されました。統計的評価の結果、臨床試験症例では85%から100%、患者支援プログラム (PSP) 症例では60%から100%の一致スコアが達成されました。文献症例については、サンプル数が限られているため厳密な統計評価は困難であるものの、モデル性能は67%から100%の範囲を示し、他の症例タイプと整合する定性的結果が確認されました。

モデルが達成した高い一致スコアは、非構造化データ源から正確かつ関連性の高い情報を抽出できる能力を示しています。これは、業務運用における具体的な効率向上として反映され得るものです。

第二の結果は、ビジネス影響評価において確認された効率向上に関するものです。ケース受付業務にLLMを導入した結果、推定39%の効率向上が確認され、1症例あたり約20分の時間短縮に相当しました。具体的には、1症例あたり平均69.4件のデータポイントが抽出され、そのうち手動修正を必要としたのは平均2.4件のみでした。

LLMの導入は単なる技術的改善にとどまらず、業務効率の向上およびPV実務における高品質な成果の確保に向けた戦略的な取り組みを意味します。

### A4. 課題および得られた知見

本概念実証 (PoC) から得られた知見は、今後の研究を支援する出発点として活用可能な五つの主要な検討事項 (Points to Consider : PtC) として整理されました。実務的な産業視点を踏まえつつ、本分野における科学的知見と関連付けながら、著者らは革新的技術の導入に関する考察を行っており、ここで共有される経験は予備的なものであるものの、同領域に取り組む関係者に有用な示唆を与えるものです。

1. 投資対効果 (RoI) はビジネスの観点で測定可能である必要があります。

LLM導入におけるRoIは定量化可能でなければなりません。本PoCでは最大39%の効率向上が確認されており、これは財務的利益およびチーム生産性向上につながる可能性があります。

2. SMEの早期関与はRoIを向上させます。

専門家 (SME) を早期に関与させることは、モデル性能の最適化、業務理解の向上および効果的なプロンプト設計を可能にするうえで重要です。これにより、LLMの限界に対応しながら信頼性および資源効率が向上し、結果としてRoIの改善につながります。

3. 規制上の不確実性は依然として大きな障壁です。

AI技術におけるGxP適合を達成するうえで、規制の不確実性は重要な課題となります。EMAや米国FDAなど主要規制当局による規制枠組みが進化し続けているため、

リアルワールドでの活用の妥当性および正確性を確保するには、先行的なリスク管理および実務的かつ解決志向のアプローチが必要となります。

4. システム統合は運用環境に即して検討する必要があります。

LLMのシステム統合は、既存システムの制約および利用者要件を考慮した運用環境の中で検討されなければなりません。意味のある研究結果を得るためには、適切なプロンプト戦略ならびに専用の前処理および後処理が重要であり、既存の安全性関連ソリューションへの効果的な組み込みが求められます。

5. 組織的準備は技術要素にとどまりません。

新技術導入に向けた組織的準備には、技術的能力だけでなく、人間の関与、十分な信頼および強固な監督体制が必要です。これらは、運用チームとの早期連携、意識改革、教育訓練および業務プロセスの整備を通じて促進され、導入阻害要因を低減し効果的な研究実施を可能にします。

これらの主要な検討事項を効果的に実装するためには、リスクに基づくアプローチが戦略的枠組みとして有効です。すなわち、リスクに基づくアプローチは、不正確性が患者安全へ及ぼし得る影響の評価、偽陽性に対する有効な低減策の策定および規制適合の確保、さらにLLM性能の継続的監視および評価、ならびに既存システムへの本技術統合の最適化を可能にします。

## A5. ガバナンスフレームワークへの適合

表11: ユースケースA: ガバナンスフレームワークとの整合性 (詳細)

出典：CIOMS Working Group XIV

原則	活動内容	SPEC	DEV	PreD	PstD	RU
<b>リスクに基づくアプローチ</b>	ケース処理におけるデータ抽出へ大規模言語モデルを導入する際に関連するリスクを分類するためには、不正確性が患者安全および医薬品安全性監視の結果へ与える潜在的影響を評価することが重要です。さらに、偽陽性に対する有効な低減策を策定し、GDPRおよびHIPAAを含む規制要件への適合を確保するため、関係者の関与が求められます。LLM性能の継続的な監視および評価に加え、利用者への十分な教育訓練は、リスク管理および既存の医薬品安全性監視システムへの本技術統合の最適化に不可欠です。	A	A	N/A	N/A	N/A
<b>ヒューマン・オーバーサイト</b>	利用者にとって操作しやすいインターフェースおよび原資料への参照機能を含む、ヒューマン・オーバーサイトを支援する専用機能を実装しました。抽出結果のすべてについて人間による品質管理（QC）を100%実施することで、結果の頑健性を確保しています。	A	A	N/A	N/A	N/A

原則	活動内容	SPEC	DEV	PreD	PstD	RU
<b>妥当性と頑健性</b>	継続学習は適用しておらず、モデルは固定状態で使用されています。新バージョンのリリース時には、十分に広範なテストセットを用いた品質保証が実施されます。	A	A	N/A	N/A	N/A
<b>透明性</b>	モデル性能は一致スコアにより測定されています。不一致の修正結果は、定期的にプロンプト戦略を改善するためのフィードバックとして活用されます。	A	A	N/A	N/A	N/A
<b>データプライバシー</b>	本サービスはプライベートクラウド上に構築されています。アクセスはプロジェクトチームメンバーのみに限定されています。個人識別可能情報は、実際のデータ抽出工程の前に匿名化処理が実施されています。	A	A	N/A	N/A	N/A
<b>公正性と公平性</b>	該当なし。本アプリケーションはデータ統合または意思決定支援を提供するものではありません。データ抽出結果の1対1一致は、人間による品質管理によって検証されています。	N/A	N/A	N/A	N/A	N/A
<b>ガバナンスとアカウントビリティ</b>	LLMモデルは、データ抽出試験のためにベンダー環境内で維持される専用プロンプト戦略を用いて運用されています。モデルはOpenAIにより提供され、複数の大規模言語モデルを基盤として稼働しています。ケース受付および処理チームが説明責任を引き継ぎ、100%の人間による品質管理プロセスを実施します。最終的な説明責任は販売承認保有者者(MAH)にあります。	A	A	N/A	N/A	N/A

## 略語

SPEC (Collection of specifications)：仕様および要求事項の収集

DEV (Development and change management)：開発および変更管理

PreD (Pre-deployment & post-change sign-off)：導入前および変更後承認

PstD (Post-deployment & post-change hyper-care)：導入後および変更後ハイパーケア

RU(Routine Use)：通常運用

A (Applicable)：適用あり

NA (Not Applicable)：該当なし

## ユースケース B: ケースの重複除外

出典：<sup>2</sup>

PV領域：ICSR Processing

### B1. ビジネス上の背景および課題

有害事象報告システム（AERS）は、医療製品の使用に関連する安全性シグナルの特定および評価を支援するため、医薬品安全性監視（PV）において重要な役割を担います。安全性監視における専門家レビューには、データマイニングや症例シリーズ解析など複数の工程が含まれますが、これらはAERSデータの品質に大きく依存します。代表的な品質上の問題の一つが重複報告であり、同一製品に対する同一患者の同一有害事象経験を複数の報告が記述している場合を指します。重複報告は、実際の有害事象数の誤認につながり、安全性評価者の作業負担を増加させるだけでなく、製品と事象の関連性を実際より弱くまたは強く見せてしまうことにより、誤った安全性シグナルの検出または見逃しを引き起こす可能性があります。

### B2. ソリューション

医薬品および生物学的製剤に関するAERSを運用し、2,800万件を超える過去報告および1日平均8,000件の新規報告を管理している規制当局は、過去および新規のすべての有害事象報告を効率的に重複排除する解決策を求めました。同当局は学術機関と連携し、最新技術（主として自然言語処理、ネットワーク解析およびクラウドコンピューティング）を活用し、構造化データおよび自由記述テキストを組み合わせた重複排除パイプラインを開発しました。本パイプラインでは、まず人口統計情報およびその他の特徴に基づく類似性の最低要件を設定し、報告ペアを絞り込む初期フィルタリングを実施します。続いて、重複検出アルゴリズムを実装したペアワイズ型〔訳注：2件1組処理〕の合理化された処理プログラムにより、条件を満たすすべての報告ペアについて確率的比較を行い、確率重みスコアおよび第二の構成要素スコアの二つを算出し、報告間の類似度を評価します。第三段階では、専用分析で設定された検証済み閾値を超える報告ペアを、潜在的な重複報告のネットワーク（別名：グループ）として統合し、さらに実際の重複からなる密接に関連したコミュニティ（別名：グループ）へ分割します。最終段階では、参照症例選択機能が複数のパラメータに基づいて各重複グループ内で最も代表性の高い報告を特定し、残りの報告は重複としてフラグ付けされ、以後のデータマイニング計算から除外されます。症例シリーズ解析を支援する既存の意思決定支援ツールにより重複報告グループの評価および参照症例の確認が可能となり、医療レビューアが常に関与する体制が維持されています。

### B3. 結果

初期研究では、重複検出アルゴリズムを市販後報告データセット二種類（ワクチン製品報告および生物学的製剤報告）に適用した結果、既知の重複ペアの77%および13%をそれぞれ特定し、いずれの場合もほぼ完全な適合率（それぞれ95%および100%）を達成しました。<sup>3</sup> その後の改良により性能はさらに向上し、医薬品有害事象報告を用いた新たな評価では、F値（F-measure）が0.9を超える重複検出性能が確認されました。新たな評価に参加した医療レビューアはアルゴリズムに高い信頼を示し、実務での利用に関心を表明しました。<sup>4</sup> 続いて、医療レビューアは体系的な手順により、重複ラベル付き2,300件の報告からなるゴールドスタンダードデータセットを作成し、新たに構築された重複排除パイプラインの検証に使用しました。その後、規制当局で従来使用されていた重複排除手法との比較が行われ、本パイプラインはそれらを上回る性能を示し、ETLプロセス

(抽出extract・変換transform・ロードload)において過去データおよび新規受信データのすべてを処理する用途として承認されました。2025年7月30日時点で、本パイプラインはAWS(Amazon Web Services)環境上に導入され、当局のAERSと密接に統合されており、3,000万件を超える過去報告を処理するとともに、1日平均8,000件の新規報告の重複排除を継続的に実施しています。

#### B4. 課題および得られた知見

本重複排除パイプラインは、複数年にわたる検討を経て開発され、所望の性能を達成し本ソリューションを本番環境へ移行するために多様な人的資源およびその他の資源が投入されました。既存手法を上回る性能を有し、かつ検証済みで透明性の高いAIベースのソリューションが、その基盤コードとともに規制当局へ無償提供されたことにより、重複排除結果の活用および既存システムへの統合に関する新たな機会が生まれました。これにより利点が最大化されるとともに、専用ツールに伴う多大なコストの削減が可能となりました。

本重複排除は日々大量のAERSデータを自動処理しており、その出力結果を人間がすべて確認および検証することはできません。そのため、これまでのすべての評価段階で示された性能が継続的に高水準で維持されていることを保証する戦略の構築が不可欠です。現時点では十分に確立された品質保証(QA)計画が存在しておらず、利用者コミュニティにおける信頼性向上に向けた重要な課題となっています。一方で、前述および図8に示された意思決定支援ツールにより、重複報告グループのレビューおよび報告された有害事象を最も代表する症例、すなわち参照症例の確認が可能となっています。このプロセスは症例シリーズ解析の場面で実施されるものであり、すべてのデータに対して適用することはできませんが、より小規模なデータセットに対する限定的評価を通じてQA計画を支援する可能性があります。この方法により、性能が同水準で維持されているか、また修正戦略が必要かどうかを判断できます。規制当局はこれらの点を慎重に検討しており、QA戦略の一環として、本手法またはその他の仕組みを用いた定期監査の実施を計画しています。

## B5. ガバナンスフレームワークへの適合

表12: ユースケースB: ガバナンスフレームワークとの整合性 (詳細)

出典: CIOMS Working Group XIV

原則	活動内容	SPEC	DEV	PreD	PstD	RU
<b>リスクに基づくアプローチ</b>	リスクに基づくアプローチについては、特に重複報告の見逃しおよび偽陽性に関して広範に検討されました。本パイプラインを人間が最終的な判断を行う意思決定支援システム内に実装することで、症例シリーズ解析におけるリスクは排除されると判断されています。今後の課題として、データマイニング計算およびシグナル検出における潜在的ノイズに関するリスクの認識が挙げられます。この部分はまだ十分に確立されておらず、現在完全に実装されている症例シリーズ解析用途ではなく、主としてデータマイニング計算における通常運用に影響します。	A	A	A	A	A
<b>ヒューマン・オーバーサイト</b>	開発段階において、人間の専門家がソフトウェアエンジニアへ積極的にフィードバックを提供し、重複排除結果を評価することでパイプラインの改良および検証を行いました。通常運用においては、既存の意思決定支援ツールを用いて症例シリーズ解析を実施する際に、専門家が参照症例の選択を確認または修正できます。本重複排除パイプラインの通常運用中には、導入前段階で示された性能が継続的に維持されていることを確認するため、定期監査が不可欠です。	A	A	A	A	A
<b>妥当性と頑健性</b>	本重複排除パイプラインは、期待される性能を満たし意図された目的を達成することを確認するために評価および検証が実施されています。ただし、重複排除後データがデータマイニング計算および潜在的な安全性シグナル検出に与える影響については、主要な用途の一つであるにもかかわらず、まだ検討されていません。	A	A	A	A	A
<b>透明性</b>	複数の学術論文、技術報告書およびその他の文書において、本パイプラインおよび安全性レビューの協力のものと実施されたすべての評価結果が記述されています。	A	A	A	A	A

原則	活動内容	SPEC	DEV	PreD	PstD	RU
<b>データプライバシー</b>	すべての処理が安全なクラウド環境内で実施されているため、本原則に完全に適合しています。	A	A	A	A	A
<b>公正性と 公平性</b>	本重複排除パイプラインは複数回の評価および検証を経ており、導入後段階においても継続的に監視されています。本報告書作成時点では、本パイプラインは本番環境へ完全移行され、通常運用として使用されていますが、本プロセスが完全に完了していないため、部分的整合として評価されています。	A	A	A	A	A
<b>ガバナンスと アカウントビ リティ</b>	システム管理者は重複排除パイプラインおよび意思決定支援ツールにおける出力利用を完全に管理し、継続的な監視を実施しています。また、重複排除結果をデータマイニング計算へ組み込む計画も策定されています。開発、導入前および導入後の各段階において明確な役割分担が定義されており、請負業者がパイプラインの構築および規制当局環境への統合を主導し、最終利用者およびその他の関係者がこれを支援しました。ただし、通常運用段階における役割分担はまだ完全には確立されていません。	A	A	A	A	A

## 略語

SPEC (Collection of specifications)：仕様および要求事項の収集

DEV (Development and change management)：開発および変更管理

PreD (Pre-deployment & post-change sign-off)：導入前および変更後承認

PstD (Post-deployment & post-change hyper-care)：導入後および変更後ハイパーケア

RU(Routine Use)：通常運用

A (Applicable)：適用あり

NA (Not Applicable)：該当なし

# ユースケース C: 人工知能翻訳アシスタント

出典：<sup>5</sup>

PV領域：ICSR Processing

## C1. ビジネス上の背景および課題

ICSR処理は、世界各国の市場から情報を収集し、電子データベースシステムまたはワークフローへ取り込むことから開始されます。通常、この過程では多言語への対応が必要となり、その後のグローバル機能による処理のため英語への翻訳が求められます。翻訳は極めて重要な工程であり、翻訳上の誤りは後続工程における誤解や誤った結論につながる可能性があります。さらに、手動翻訳には時間と労力を要し、すべての市場および言語を熟練翻訳者でカバーすることは課題となります。

本事例では、製薬企業がグローバルなケース受付および翻訳プロセスの統合と効率化を目的としてベンダーと協働しました。当該ベンダーは、複数の各国組織および地域ベンダーによる分散型体制を置き換える形で、欧州およびアジアに二つのハブを設置し、32か国16言語で対応可能な運用体制を構築しました。さらに生産性向上を目的として、翻訳プロセスの自動化が求められました。

## C2. ソリューション

外国語による有害事象報告の処理においては、原資料を現地言語から英語へ正確に翻訳する作業が全体作業量の約半分を占めており、英語による集中管理および規制当局への提出を可能にしていました。製薬企業とベンダーは、機械学習専門家およびPV担当者から構成される共同プロジェクトチームを組織し、市販技術を基盤としたAI翻訳アシスタントの試験導入を実施しました。チームは学習データ（原文および人間による修正済翻訳）を保存するためのプライベートクラウド環境を構築し、原文入力および翻訳結果の取得、（必要なら）編集を可能とするユーザーインターフェースを開発しました。

本システムは、利用者による修正内容を自動的に保存および解析し、AI翻訳アシスタントの初回翻訳品質を向上させるための学習反復に活用します。翻訳の正確性を常に保証するため、すべての翻訳について人間による100%品質管理（QC）が実施されました。本ソリューションは、定期的到手動修正内容を翻訳モデルへ自動統合することにより継続学習を可能としています。モデル更新のたびに、関連する品質指標（BLEUスコア、後述）が再計算されます。現在に至るまで、ヒューマン・イン・ザ・ループ（HITL：Human-in-the-Loop）による100%QCが維持されています。

## C3. 結果

翻訳品質はBLEUスコアにより評価されました。BLEUは機械翻訳テキストを評価する指標であり、機械翻訳結果と高品質な参照翻訳との類似度を0から1の数値で示します。6か月以内に、AI翻訳アシスタントは人間による翻訳者と同等の品質（すなわちBLEUスコア0.6以上）を達成しました。<sup>6</sup>

最初の対象言語（ポルトガル語）におけるパイロット結果では、翻訳作業量が約30%削減されました。その結果、本ソリューションはさらに5言語（中国語、オランダ語、フランス語、ドイツ語およびスペイン語）へ拡張されました。製薬企業およびベンダーチームは、ソリューション品質を監視するため、BLEUスコアを共同かつ継続的に評価しています。

AIモデルの改善は症例数に依存しており、QCチームが提供する修正済翻訳サンプルの蓄積がモデル性能向上に寄与します。サンプル数が増えるほどモデルは改善され、最終的にはチームの作業負担が軽減され、より多くの症例をより迅速かつ一貫性をもって処理できるようになります。

#### C4. 課題および得られた知見

2021年にAI翻訳アシスタントが導入されて以降、翻訳品質はさらに向上し、現在では人による修正を必要とする出力は10%未満となっています。チームでは、全面的な人間による品質確認から、言語ごとに必要に応じてサンプル数を調整するサンプルベース監視への移行を検討しています。

2022年以降の生成AI（GenAI）の発展により、これらのツールは、構造化データの抽出情報翻訳、症例要約の作成、非英語地域向け報告書の翻訳など、複数の機能を単一プラットフォーム内で提供できるようになりました。現在、ICSR管理向け商用アプリケーションへGenAIを導入する複数のプロジェクトが進行しています。これらのプラットフォームは、ICSR管理における効率性を大幅に向上させる可能性を有しています。現時点では、高品質を確保するためにヒューマン・イン・ザ・ループ（HITL, Human-in-the-Loop）が引き続き重要な役割を担います。

#### C4. ガバナンスフレームワークへの適合

表13: ユースケースC: ガバナンスフレームワークとの整合性 (詳細)

出典：CIOMS Working Group XIV

原則	活動内容	SPEC	DEV	PreD	PstD	RU
<b>リスクに基づくアプローチ</b>	受領情報の翻訳には誤りが生じるリスクがあり、それが後続工程における誤った結論または評価につながる可能性があります。そのため、リスクに基づくアプローチが徹底して適用され、継続的かつ完全なヒューマン・オーバーサイトを確保するための措置が講じられています。	A	A	A	A	A
<b>ヒューマン・オーバーサイト</b>	ヒューマン・オーバーサイトを確保するため、導入当初からベンダー翻訳者による翻訳テキストの100%の人間による品質管理（QC）が実施されています。各言語についてBLEUスコアが定期的に測定され、全体的な性能変化の把握に利用されています。	A	A	A	A	A
<b>妥当性と頑健性</b>	本システムは、ベンダーの標準バリデーション手法に従って実装されています。100%の人間による品質管理により、すべての翻訳結果の妥当性が確認されています。翻訳アシスタントに不具合が生じた場合には、直ちに検出および修正が可能です。	A	A	A	A	A

原則	活動内容	SPEC	DEV	PreD	PstD	RU
<b>透明性</b>	翻訳性能の透明性は、すべての翻訳結果および人間による修正内容がシステム上で追跡されることにより確保されています。これらの修正は、モデル改善のため定期的に活用されます。	A	A	A	A	A
<b>データプライバシー</b>	本サービスはプライベートクラウド上に構築されています。アクセスはプロジェクトチームメンバーのみに限定されています。個人識別可能情報は実際の翻訳処理前に匿名化されています。原資料は、初期情報を受領し、必要に応じて初回報告者へフォローアップを行う可能性のある現地チームのみがアクセス可能な状態で保持されます。	A	A	A	A	A
<b>公正性と 衡平性</b>	本アプリケーションはデータ解釈、統合または意思決定支援を提供するものではありません。AI翻訳アシスタントが特定の語句や表現（例：口語表現）を認識できない場合には、品質管理工程において人間による翻訳者が対応します。翻訳結果の1対1一致は常に人間による品質管理で検証されます。	A	A	A	A	A
<b>ガバナンスと アカウントビ リテイ</b>	翻訳アシスタントはベンダー企業が所有する独立したツールです。そのため、通常のライフサイクルガバナンスはベンダーにより実施され、製薬企業からの要請に応じて提供されます。これには、学習進捗に基づくモデル更新などが含まれます。翻訳実行の責任はベンダーにあります。最終的な説明責任は製薬企業に残ります。そのため、ベンダーによる100%の人間による品質管理に加え、製薬企業は翻訳を含むケース受付全体の結果について、定義されたサンプル品質管理を実施しています。	A	A	A	A	A

## 略語

SPEC (Collection of specifications)：仕様および要求事項の収集

DEV (Development and change management)：開発および変更管理

PreD (Pre-deployment & post-change sign-off)：導入前および変更後承認

PstD (Post-deployment & post-change hyper-care)：導入後および変更後ハイパーケア

RU (Routine Use)：通常運用

A (Applicable)：適用あり

NA (Not Applicable)：該当なし

# ユースケース D: コンテキスト対応構造化照会言語 (Structured Query Language: SQL) のための大規模言語モデル

出典: 7

PV領域: Safety analysis

## D1. ビジネス上の背景および課題

安全性評価担当者は、安全性データベースからSQLを用いてデータ抽出や安全性クエリを作成する際、技術チームへ依存することが多く、その結果として評価の遅延が生じる場合があります。本取り組みの目的は、ビジネスコンテキスト文書を活用し、大規模言語モデル (LLM) により自然言語クエリ (NLQ) を構造化照会言語 (SQL) クエリへ変換することで、PVデータベースからの情報取得精度を向上させることでした。

## D2. ソリューション

本研究では、OpenAIのGPT-4モデルのサンドボックス版を、ビジネスコンテキスト文書で強化したRAG (Retrieval-Augmented Generation) フレームワーク内で使用し、自然言語クエリを実行可能なSQLクエリへ変換しました。研究は三つの段階で実施され、クエリの複雑性を変化させながら、ビジネスコンテキスト文書の有無によるLLM性能を評価しました。

RAGフレームワークはその検索機構を用いてビジネスコンテキスト文書から関連情報を取得し、自然言語クエリ (NLQ) からSQLクエリへの変換を促進します。その拡張された文脈データをGPT-4モデルへの入力として提供することで、特定のデータベーススキーマおよび運用要件に整合したSQLクエリ生成を可能にします。

## D3. 結果

結果は、三つの実験段階すべてにおいてクエリ生成精度の大幅な改善を示しました。フェーズ1では、データベーススキーマのみを使用した場合、LLMの合格率は8.3%にとどまり、78.3%が有効なSQLクエリを生成できず、コンテキスト情報なしで正確なSQL生成を行う困難さが示されました。フェーズ2では、ビジネスコンテキスト文書を追加することで合格率は78.3%へ向上し、フェーズ1と比較して統計的に有意な改善 (P値: 0.0006) が確認されました。フェーズ3では、ビジネスコンテキスト文書を用いずにスキーマを限定した条件で評価を行い、フェーズ1と比較して失敗率は78%から50%へ低減したもののフェーズ2の性能には到達しませんでした。

本手法は、非技術者が複雑なデータクエリを実行できるよう支援する補助的手法でありPVデータ解析および報告の迅速性向上に寄与する可能性があります。

## D4. 課題および得られた知見

本研究は、LLMを用いたPVデータベース向けSQLクエリ生成の自動化における課題を明らかにしました。一つの制約は高複雑度クエリへの対応の難しさであり、データベース関係が複雑で利用者の意図が曖昧な場合、LLMは正確なSQLコード生成に苦慮しました。この課題は特にフェーズ1で顕著であり、コンテキスト情報の欠如により失敗率78.3%という結果となりました。フェーズ2ではビジネスコンテキスト文書の導入により性能は大きく改善し、合格率78.3%を達成しました。さらに、実装上の考慮事項として、ビジネスコンテキスト文書の作成および維持にはドメイン専門家の関与が必要であり、データ

ベーススキーマの更新がその有用性へ影響を与える可能性がある点が挙げられます。

## D5. ガバナンスフレームワークへの適合

表14: ユースケースD: ガバナンスフレームワークとの整合性 (詳細)

出典: CIOMS Working Group XIV

原則	活動内容	SPEC	DEV	PreD	PstD	RU
<b>リスクに基づくアプローチ</b>	本研究では、安全性データベースからデータを取得するために自然言語を用いてSQLクエリを生成する利用可能性を示すことを目的としています。リスクに基づくアプローチでは、その信頼できる利用を確保するために、実装可能性、必要な統制およびプロセスを考慮する必要があります。	A	A	N/A	N/A	N/A
<b>ヒューマン・オーバーサイト</b>	PoCにおいては、人間の専門家が参照基準と照合しながら出力結果の妥当性をレビューしました。製品環境においては、抽出されたデータが正しいかを確認することを含め、出力の頑健性を確保するためのヒューマン・オーバーサイトの方法を検討する必要があります。さらに、データベーススキーマが変更される可能性があるため、一定期間ごとの性能監視についても検討が必要です。	A	A	N/A	N/A	N/A
<b>妥当性と頑健性</b>	本ツールは、精査された参照基準に対して評価されています。PoCを超えて本番環境で使用する場合には、利用者要件に基づき出力結果が正確であることを保証するなど、汎用性についての検討が必要となります。	A	A	N/A	N/A	N/A
<b>透明性</b>	GPTモデル自体の透明性に加え、RAGおよびコンテキスト特化文書の利用により、出力を生成するためのデータ処理パイプラインの透明性が確保されています。	A	A	N/A	N/A	N/A
<b>データプライバシー</b>	本ツールはSQL出力生成において個別患者データを使用しない補助ツールです。	N/A	N/A	N/A	N/A	N/A
<b>公正性と公平性</b>	本ツールは個別患者データを使用しない補助ツールです。	N/A	N/A	N/A	N/A	N/A

原則	活動内容	SPEC	DEV	PreD	PstD	RU
<b>ガバナンスと アカウンタビ リティ</b>	PoC段階においては、本手法に関する説明責任は開発者にあります。しかし、本手法が本番環境へ統合された場合には、説明責任は人間の主題専門家(SME: Subject matter expert)へ移行します。PoCにおけるガバナンスは科学的完全性の原則が遵守されることを保証するものであり、将来的な製品利用におけるガバナンスでは、本ツールがPVシステム全体および品質マネジメントシステム (QMS) 内でのように位置付けられるかを対象とする必要があります。	A	A	N/A	N/A	N/A

## 略語

SPEC (Collection of specifications) : 仕様および要求事項の収集

DEV (Development and change management) : 開発および変更管理

PreD (Pre-deployment & post-change sign-off) : 導入前および変更後承認

PstD (Post-deployment & post-change hyper-care) : 導入後および変更後ハイパーケア

RU(Routine Use) : 通常運用

A (Applicable) : 適用あり

NA (Not Applicable) : 該当なし

# ユースケース E: 薬物有害反応の因果関係評価

出典：<sup>8</sup>

PV領域：Causality Assessment

## E1. ビジネス上の背景および課題

有害事象と患者の薬剤曝露との因果関係を評価することは、PVプロセスにおける重要な工程であり、各ICSRにおける迅速報告要件の判定に直接影響します。因果関係評価は時間を要する作業であり、医学専門家が症例情報を外部情報（例：添付文書、科学文献、薬剤の作用機序、疾患症状）と照合しながら手動でレビューを行う必要があります。レビュー対象となる有害事象数が増加する中で、機械学習（ML）を活用し、因果関係評価を予測することで医学専門家を支援するソリューションを構築する機会が生まれています。

## E2. ソリューション

本論文の著者は、製薬企業の安全性データベースに収載された自発報告以外の収集症例（solicited cases）から、薬剤—事象組合せの因果関係評価に関連する複数のデータ属性を用いてモデリング用特徴量セットを作成しました。さらに、外部データおよび他の内部情報源から生成された特徴量を追加しました。その結果得られた学習データスキーマ（下表15参照）は、因果関係評価プロセスに関連する包括的な特徴群を提供するものとして選定されました。

表15: ユースケースE: データのモデリング

出典：Modified from Cherkas Y, et al, 2022<sup>9</sup> Table reproduced with permission

モデリングデータ		
症例レベルデータ		外部取得データ
因果関係ラベル	既往歴除外情報	不均衡性
再投与(Rechallenge)	薬剤除外情報	解剖学的治療分類および器官別分類
ラベル情報の有無		時間的關係
報告者による因果関係評価		時間的整合性

並行して、経験豊富な医薬品安全性担当医師との協議を通じて、別個の意思決定支援ツール（CASCADE）が開発および検証されました。他の因果関係評価アルゴリズムと比較して透明性および解釈可能性が高いことから、決定木構造（decision tree structure）が採用されました。この高い透明性および解釈可能性により、評価根拠を明確に記述することが可能となります（例：「本症例は、(a) 当該事象が添付文書に記載されていること、(b) 事象に合理的な時間的關係が存在すること等により、因果関係ありと判断される」）。

この決定木に関する検討は、その後の予測モデル構築の基盤となり、寄与因子および最終的なベイジアンネットワークモデルの構造設計に反映されました。

著者らが本モデルを選択した理由には、複数の情報源と専門家知識を統合できる点、透明性および解釈可能性を有する点、不確実性が大きい状況において因果依存関係を含む複雑な枠組みをモデル化できる能力が挙げられます。モデル学習には5万件のアノテーション付き症例データセットが使用され、別途2万件のテストデータセットが用いられました。学習データおよびテストデータはいずれも幅広い薬剤分類および事象カテゴリーを含んでいます。すべての症例は過去に医学専門家によって評価済みであり、因果関係評価手法が一貫していた期間から抽出されました。

### E3. 結果

本モデルは、グローバルイントロスペクションに基づく臨床判断と比較して、薬剤—事象組合せの因果関係評価予測において高い性能を示しました（感度0.900、陽性的中率 [PPV] 0.778）。著者らは同一の学習データを用いてトポロジー学習型ベイジアンネットワークモデルについても検討しましたが、この学習型トポロジーモデルはCASCADEに基づくモデルと比較して性能が劣ることが確認されました。

### E4. 課題および得られた知見

データの入手可能性は複数の課題を伴う一方で、新たな特徴量の追加やモデル性能のさらなる検証を通じてモデル改善の機会も提供します。

追加情報源から十分に注釈付けされた因果関係データが不足していたため、内部データセットに含まれない薬剤—事象組合せに対するモデル性能の検証は限定的となりました。公開参照データセットの構築は困難が予想されるものの、こうしたモデルの検証および他手法との性能比較を、より広範な薬剤および事象にわたって実施できる可能性をもたらします。

本研究では、薬剤曝露から事象発現までの妥当な発現時間を推定するため、関連する臨床試験症例の限定的なデータセットが使用されました。電子健康記録（EHR）など、より包括的なデータセットを利用できれば、このモデル設計要素の改善が期待されます。また、発現時間に関する特徴量へ薬剤作用機序データを組み込むことも、モデル性能向上に寄与する可能性があります。

薬剤ラベル情報へのアクセスは、症例中の薬剤または同一薬効群の薬剤において事象が添付文書に記載されているかを識別する特徴量の追加を可能にします。同様に、医学的狀態および薬剤適応に関するデータを組み込むことで、報告された反応が既存疾患の症状であるか、あるいは併用薬の適応に関連するものであるかといった交絡因子の特定に役立つ可能性があります。

薬剤—事象組合せにおける因果関係評価のばらつきは広く知られており、本活動を支援するモデル設計において透明性を確保する上で課題となります。検証済みの決定木ツール（CASCADE）の開発により、構造化され一貫性かつ透明性のあるアプローチが提供され、最終モデルの構造設計に重要な示唆を与えるとともに、機械学習モデルへ臨床的専門知識を統合する価値が示されました。ただし、モデルの解釈可能性は依然として解決すべき課題として残されています。

## E5. ガバナンスフレームワークへの適合

表16: ユースケースE: ガバナンスフレームワークとの整合性 (詳細)

出典: CIOMS Working Group XIV

原則	活動内容	SPEC	DEV	PreD	PstD	RU
<b>リスクに基づくアプローチ</b>	本研究の対象範囲および構築されたモデルは、市販後に収集された症例に限定されています。これら症例の因果関係評価を自動化することは、PVシステムへのリスクおよび影響が比較的低いと判断されました。	A	A	N/A	N/A	N/A
<b>ヒューマン・オーバーサイト</b>	医薬品安全性担当医師および専門家がデータレビューおよびモデル開発活動に関与し、モデルが意図された目的に適合することを確保しました。ただし、将来的な本番運用におけるヒューマン・オーバーサイトを支援する品質マネジメント枠組みの構築については言及されていません。	A	A	N/A	N/A	N/A
<b>妥当性と頑健性</b>	本論文ではユースケースおよび適用領域が明確に記述されています。本研究で使用されたデータは、幅広い製品および事象カテゴリーにおいて因果関係評価手法が一貫して適用されていた特定の期間に限定されており、結果モデルの信頼性向上が図られました。モデル学習およびテストの方法は詳細に記載されており、性能評価手法についても説明されています。さらに、薬剤—事象因果関係評価に関する公開参照データセットの利用可能性など、モデルの妥当性向上および頑健性強化に向けた今後の検討領域も提示されています。	A	A	N/A	N/A	N/A
<b>透明性</b>	本論文全体を通じて透明性が重視されています。モデルの意図された用途および設計に関する情報が提示されています。また、因果関係評価結果の根拠を明確に示すことを目的とした決定木ツール (CASCADE) が開発され、最終モデル設計へ反映されています。データ、結果、今後の検討領域、およびPVシステムにおける適用方法についても議論されています。	A	A	N/A	N/A	N/A

原則	活動内容	SPEC	DEV	PreD	PstD	RU
<b>データプライバシー</b>	モデルの開発、学習および検証に使用されたデータは、企業の内部市販後安全性データベースに由来しており、患者または報告者の同意のもと取得され、関連するプライバシー法令および規制に従っていることが示されています。	A	A	N/A	N/A	N/A
<b>公正性と 衡平性</b>	本論文の記載内容からは、モデル開発が本指針原則に整合しているかについて判断することはできません。	N/A	N/A	N/A	N/A	N/A
<b>ガバナンスと アカウントビ リティ</b>	本指針で定義されるガバナンスとアカウントビリティに関する活動について、本論文では言及されていません。ただし著者らは、モデルが規制枠組みおよびガイドラインに準拠し続ける必要性を認識しています。また、作成された決定木CASCADEは因果関係評価支援ツールとして位置付けられており、最終的な因果関係評価の判断責任は薬剤安全性分野の専門家に残ることが示唆されています。	N/A	N/A	N/A	N/A	N/A

## 略語

SPEC (Collection of specifications)：仕様および要求事項の収集

DEV (Development and change management)：開発および変更管理

PreD (Pre-deployment & post-change sign-off)：導入前および変更後承認

PstD (Post-deployment & post-change hyper-care)：導入後および変更後ハイパーケア

RU(Routine Use)：通常運用

A (Applicable)：適用あり

NA (Not Applicable)：該当なし

# ユースケース F: シグナル検出を支援するプロセス効率

出典：<sup>10</sup>

PV領域：Signal Detection

## F1. ビジネス上の背景および課題

安全性シグナルを否定するために最も時間および資源を要する作業の一つは、不均衡な報告によりシグナルが特定された後、ICSRに報告された有害事象（AE）について代替的な原因の可能性を特定することです。これには、新たに特定された薬剤—事象組合せに対する代替原因の可能性を確認するため、併せて報された薬剤のスクリーニングが含まれます。

## F2. ソリューション

本研究は、AIベースのフレームワークを開発し、(1) 不均衡性解析における対照群の選択および(2) 偽陽性の不均衡シグナルを否定する目的で代替原因となり得る薬剤の特定を自動化することを目的としました。

条件付き推論木（conditional inference tree）を用いた対照群の自動選択および偽陽性シグナルの除外プロセスの実装は、以下のフローチャートに要約されています。

### 図 6：条件付き推論木を用いたコントロールの自動選択と偽陽性シグナルの棄却の実装をまとめたフローチャート

出典：Al-Azzawi F et al, 2023 359 Reproduced under Creative Commons Attribution-Non Commercial 4.0 International License. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.



対照群の自動選択を目的として、FAERSを用いた不均衡性解析では、ATC分類コードおよびガルクネズマブの米国添付文書（US Prescribing Information：USPI）に記載された承認適応を組み合わせた二重アプローチが採用されました。ガルクネズマブと同一の治療標的（すなわちCGRP拮抗薬）を有するすべての有効成分は、ATCコード第4階層、すなわち化学サブグループを用いて特定されました。さらに、DrugBankを用いて、ガルクネズマブと同一の承認適応を有しつつも当該化学サブグループ外に属する有効成分を対照として特定し、薬剤クラス効果によるマスキングおよび適応による交絡を回避することを目的としました。

不均衡性シグナルはさらに、条件付き推論木を用いて代替原因となる薬剤を特定する解析が実施されました。不均衡に併報された薬剤のUSPIを確認し、シグナル検証時に実施される手順を模倣する形で、当該有害事象がUSPIに記載されている薬剤を特定しました。その後、検討対象の有害事象がUSPIに記載されている併報薬剤を含む症例を除外し、これら症例が有害事象の代替原因を有すると判断したうえで、不均衡性解析を再度実施しました。

### F3. 結果

条件付き推論木を用いた結果、本フレームワークは対照群内の症例から特定された純粋な代替原因に基づき、エレヌマブでは20.00%、トピラマートでは14.29%、アミトリプチリンでは13.33%の不均衡性シグナルを除外することができました。さらに、代替原因のみでは除外できなかった不均衡性シグナルについても、手動検証が必要となるガルカネズマブ症例数が、エレヌマブ、トピラマートおよびアミトリプチリンと比較して、それぞれ15.32%、25.39%、26.41%減少すると推定されました。

著者らは、AIがシグナル検出および検証における最も時間と労力を要する工程の一部を大幅に軽減できる可能性があるとして結論付けました。本AIベースアプローチは有望な結果を示しましたが、フレームワークの妥当性確認のためには今後さらなる研究が必要とされています。

### F4. 課題および得られた知見

本研究は、FAERSデータベースにおけるAIを用いたシグナル検出自動化に関する特有の課題を明らかにしました。その一つは、不均衡性解析における対照群選択に関する明確な指針が存在しない点です。これは、不均衡性解析がしばしばデータベース全体を背景として実施されることに起因します。また、適切な対照群の選択がデータセットによっては極めて困難となる場合があり、対照選択に伴う主観性も課題となります。それでも本研究で提案されたガルカネズマブの化学サブグループ内外から対照を自動選択する手法は86%の成功率を示し、特定の状況では有望であることが示されました。

さらに、有害事象の代替原因を確認するために併報薬剤をスクリーニングする手動工程が依然として必要であり、本AIフレームワークではこの工程を完全には自動化できていない点も課題として挙げられました。条件付き推論木により併報薬剤比率の統計的差異を特定でき、偽陽性シグナルの除外を支援できるものの、製品特性概要（SmPC）の確認を自動化するための専用ツールの必要性が残されています。

加えて、本フレームワークでは、有効な対照群を特定するために必要な薬剤クラス数の明確な閾値が設定されておらず、不均衡性解析における最適な対照数も決定されていませんでした。この点はさらなる研究の必要性を示しています。また、アラート数が対照数に比例して変化することから、この多重性問題を実務上適切に管理するための体系的基準の確立は依然として困難であり、異なる薬剤およびデータベースにおける追加検証が求められています。

## F5. ガバナンスフレームワークへの適合

表17: ユースケースF: ガバナンスフレームワークとの整合性 (詳細)

出典: CIOMS Working Group XIV

原則	活動内容	SPEC	DEV	PreD	PstD	RU
<b>リスクに基づくアプローチ</b>	本研究では、シグナル検出プロセスにおける偽陽性の影響を特定および軽減するため、体系的アプローチが採用されました。シグナル検出を支援するツールを開発する際には、品質マネジメントシステム (QMS) 内において当該ツールがPVシステム全体へ与える影響を考慮し、広範な導入を支援するために必要なリスク低減策を検討する必要があります。	A	A	N/A	N/A	N/A
<b>ヒューマン・オーバーサイト</b>	本研究では、AIが生成した結果の正確なシグナル検出を確保するため、医学専門家が手動検証において重要な役割を担いました。各症例の複雑性およびばらつきを考慮し、人間によるレビューの必要性が強調されており、AI結果の解釈には当該領域専門家の関与が不可欠であることが示されています。この監督には、AIが検出した有害事象について代替原因を特定し、アルゴリズム予測と臨床的知識との整合性を確保する作業が含まれます。	A	A	N/A	N/A	N/A
<b>妥当性と頑健性</b>	本論文では、FAERSデータおよびシミュレーションを用いた検証および試験の必要性が指摘されています。統制された試験の実施および対照選択の最適化により、多様な薬剤および自発報告データベースにおける安定性および予測信頼性の向上が図られました。	A	A	N/A	N/A	N/A
<b>透明性</b>	採用された方法論および対照選択の根拠を文書化することにより透明性が確保され、ばらつきの影響への対応が図られました。	A	A	N/A	N/A	N/A
<b>データプライバシー</b>	本手法は、ラベル情報などの公開情報およびFAERSデータを使用しており、データプライバシーに関する考慮は限定的でした。	N/A	N/A	N/A	N/A	N/A

原則	活動内容	SPEC	DEV	PreD	PstD	RU
<b>公正性と 衡平性</b>	本研究では、バイアスを防止するため、対照群選択における比較可能性および包括性を確保する必要性が強調されました。不十分なデータ収集や対照設定により生じ得る不均衡に対応するため、薬剤間のデータ変動性および比較可能性を評価し、公平な代表性を確保する重要性が示されています。これは、比較可能性の複雑性とより広範な薬剤クラスの包含との均衡を図る上で重要とされています。	A	A	N/A	N/A	N/A
<b>ガバナンスと アカウンタビ リティ</b>	PoC段階においては、本手法に関する説明責任は開発者にあります。しかし、本手法が本番環境へ統合された場合には、説明責任は人間の専門分野担当者へ移行します。PoCにおけるガバナンスは科学的完全性の原則が遵守されることを保証するものであり、将来的な製品利用におけるガバナンスでは、本ツールがPVシステム全体および品質マネジメントシステム（QMS）内でどのように位置付けられるかを対象とする必要があります。	A	A	N/A	N/A	N/A

## 略語

SPEC (Collection of specifications) : 仕様および要求事項の収集

DEV (Development and change management) : 開発および変更管理

PreD (Pre-deployment & post-change sign-off) : 導入前および変更後承認

PstD (Post-deployment & post-change hyper-care) : 導入後および変更後ハイパーケア

RU(Routine Use) : 通常運用

A (Applicable) : 適用あり

NA (Not Applicable) : 該当なし

# ユースケース G: 生成AI:医薬品安全性監視評価を促進するための大規模な非構造化安全性文書リポジトリからの合成と要約

出典：Internal to CIOMS Working Group XIV member organisation

PV領域：PV document retrieval

注記：本ユースケースは、CIOMS Working Group XIVの参加組織の一つがPV関連業務プロセスの強化を目的として導入した生成AIソリューションの実装事例を示しています。世界中の多くの販売承認保有者（MAH）および規制当局（例：米国FDAのELSA）においても、業務改善の可能性を目的として生成AI技術の検討が進められている点に留意する必要があります。

本ユースケースはCIOMSによる本文書公開以前に開発および実装されたものであり、そのため記載された実務が本文書におけるベスト・プラクティスまたは推奨事項と完全には一致しない可能性があります。

## G1. ビジネス上の背景および課題

PV評価を支援するデータ整理は、安全性解析やシグナル評価から定期報告、規制当局からの安全性照会対応に至るまで、非常に時間と労力を要する作業です。これらの作業では、臨床試験報告書や当局提出資料のような大量のテキスト主体文書、過去の安全性解析、シグナル評価資料、科学文献など、膨大な非構造化安全性データの検索、取得、レビューおよび要約が必要となります。GPTモデルなどの大規模言語モデル（LLM）は多様かつ広範なリポジトリから非構造化安全性データを要約および統合する能力によりPV専門家の作業を支援できます。

## G2. ソリューション

安全性関連データを豊富に含む数十万件規模の文書を対象として、非構造化安全性データの検索、取得、レビューおよび要約を、PV専門家の要件および作業条件に適合する形で最適化するためにLLMが活用されました。

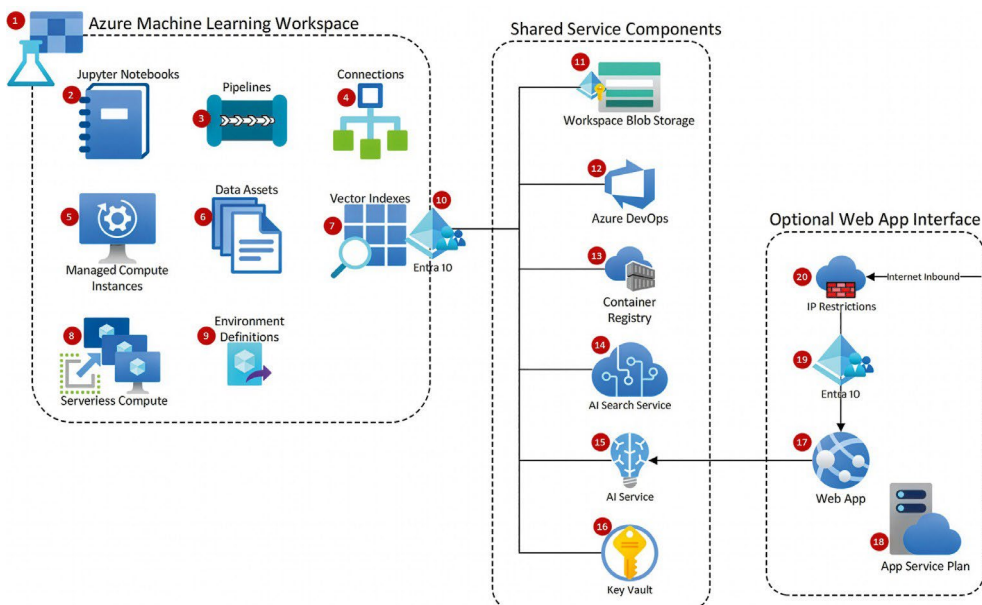
C#および.NET Frameworkを用いてカスタムAI検索エンジンが開発され、MicrosoftのSemantic Kernel SDKおよびRAG（Retrieval-Augmented Generation）パターンを活用することで、プライベートなベクトルストア内に保存された非構造化安全性データに対し、大規模言語モデルとカスタムシステム指示を用いた対話的操作を可能にしました。PV専門家はWebフロントエンドを通じて安全性関連クエリを入力し、当該クエリはコンテキスト情報（ベクトルストアから取得されたデータ）およびシステム指示とともに、プライベート環境に配置されたLLMへ送信されます。Azure OpenAIは、REST（Representational State Transfer）アプリケーションプログラミング（API）を介して、多様かつ高性能なモデル群（OpenAI Chat、OpenAIテキスト埋め込みモデル、GPT-4.1シリーズ）へのアクセスを提供し、これらが大規模な非構造化安全性データリポジトリと統合します。

OpenAIモデルの「テキスト入力・テキスト出力」インターフェースにより、PV専門家は入力プロンプトを提示し、GPT-4.1シリーズなどのモデルが対話形式でテキスト応答を生成します。また、OpenAIの埋め込みモデルを活用し、自然言語処理タスクのためにテキストを高密度ベクトル表現へ変換します。

関連情報の取得プロセスを自動化することで、PV専門家は手動によるデータ探索作業から解放され、より付加価値の高い業務へ時間を充てることが可能となります。以下の図は、Azure LLMアーキテクチャおよびデータ資産とのインターフェース構成を示しています。

図 7: 初期の人工知能アーキテクチャの概要

出典: Internal to CIOMS Working Group member organisation. Figure reproduced with permission.



会社内の関連ファイル構造に対して精緻に設計されたベクトルインデックスを構築し、展開することで、組織はLLMを活用して社内ファイルに存在する文書、データ、および情報を容易かつ効率的に探索できるようになりました。新しいモデルが公開されるたびにそれらへ移行することで、回答の正確性および有用性の向上が観察されており、これらは手動によるプロセス検証を通じて確認されています。

AIツールの性能改善および追加ユースケースの発見を目的として、定期的なフィードバックセッションが実施されています。この反復的アプローチにより、継続的な改善が可能となり、エラーの最小化が図られています。

さらに、モデルに対して効果的に質問を行うためのガイドラインも策定されており、これによりツールの実用性が一層向上しました。

### G3. 結果

本ユースケースにおけるGenAIツールは、PV担当者が数百から数千に及ぶ安全性関連文書を迅速かつ効率的に検索し、構造化された知的アウトプットを得ることを可能にするAI「リサーチアシスタント」としての有用性を示しました。

本プロジェクトの現時点では、PVユーザーはGenAIツールの機能、とりわけ安全性データの要約および検索への活用方法をより深く理解し適用するためのトレーニングを受けています。正式な定量指標は収集されていないものの、ユーザーによる品質確認（QC）が実施され、その結果はプロジェクトチーム内のGenAI開発者と共有されています。複数のPVユーザーによるGenAIツールの利用は、ツールの正確性および性能評価において重要な役割を果たしました。手動検証プロセスにより、ユーザーは本ツールが適切な検

索結果を提示し、正しい情報を取得し、さらに原資料を正確に要約できていることを確認しました。これらのフィードバックに基づき、GenAIツールは、より効率的かつ高速な応答を提供できるよう改良が進められており、安全性データ参照元をダウンロード可能な形式で提供する機能も追加されています。また、より正確で関連性の高い回答を支援するため、明確な指示文およびコンテキストプロンプトの統合も行われています。

将来的には、本GenAIの適用範囲を拡張する計画があり、シグナル検出目的の安全性文献の抽出および要約、ならびに症例報告関連アウトプットへの活用が検討されています。さらに、集積安全性報告書の要約情報作成、監査および査察対応支援、またはベネフィット・リスク評価関連業務への活用の可能性も想定されています。

例えば、「EMA迅速報告」に関する質問に対して、本GenAIツールは膨大な文書リポジトリから適切な文書を即座に検索・特定し、迅速報告要件を表形式の要約として抽出するとともに、参照元文書も提示することができ、症例処理および安全性データベース設定を支援しました。

最終的に、大規模文書リポジトリからAIを活用した文書検索および要約を行うことにより、PV専門家が安全性情報に関する重要な医学的・科学的評価をより効率的に実施できるようになり、製品の安全性および有効性の向上に寄与する可能性があります。

#### G4. 課題および得られた教訓

OpenAI活用における主な課題として、コストが高額かつ予測困難である点、必要なAI専門知識の不足、ならびにデータプライバシーおよびセキュリティに関する懸念が挙げられました。これらの課題が特定された後、当該企業はAPIコストを監視・管理するための手法を迅速に整備し、プロンプト設計の最適化、インデックス化の導入、および潜在的に機微な情報を保護するためのデータセキュリティ・プライバシープロトコルを確立しました。

GenAIプロジェクトを実装する際には、プロジェクト開始時点でのコスト管理およびコスト上限設定 (cost capping) を行い、予算を統制することが強く推奨されます。また、プロジェクト初期段階からAI専門家またはベンダーが関与することで開発を円滑に進めることが可能となり、さらにインデックスシステムの構築は、ユーザーが目的に合致した質問を効率的に行ううえで有効であることが示されました。

プログラムのさらなる発展の一環として、ユーザーによる品質評価を体系的に収集・評価し、フィードバックセッションを継続的に実施することが想定されています。これにより、時間の経過とともに開発者側に知識および経験の蓄積が形成され、GenAIツールの継続的な改善および高度化が可能となります。

#### G5. ガバナンスフレームワークへの適合

GenAIプロジェクトは、開発段階および導入前段階 (pre-deployment) において、適用範囲を限定した形で慎重に設計・管理され、表 18 に示される適用可能な指針原則との整合性が確保されました。

今後、本プロジェクトが本番運用へ移行し、利用範囲および適用範囲が拡大するに伴い、これら原則との密接な整合性を維持するための慎重な検討が行われる予定です。

そのため、本ユースケースにおけるコンプライアンスは「高い整合性 (closely aligned)」として評価されており、AIソリューションが適応性および拡張性を備えながら医薬品安全性監視業務を遂行できるという信頼性の基盤を形成しているとされています。これらの内容は表 18 においてさらに詳細に説明されています。

表18: ユースケースG: ガバナンスフレームワークとの整合性 (詳細)

出典: CIOMS Working Group XIV

原則	活動内容	SPEC	DEV	PreD	PstD	RU
<b>リスクに基づくアプローチ</b>	本ユースケースでは、GenAIツールは限定された範囲で開発され、安全性およびPV業務への影響を考慮しながら段階的に導入されました。リスク評価は、文書検索・要約結果の正確性およびPV判断への影響可能性を考慮して実施され、段階的拡張を前提とした管理が行われました。	A	A	A	N/A	N/A
<b>ヒューマン・オーバーサイト</b>	PV専門家による手動検証（QC）が実施され、検索結果および要約内容の妥当性が確認されました。ユーザーからのフィードバックは継続的改善プロセスに組み込まれています。	A	A	A	N/A	N/A
<b>妥当性と頑健性</b>	ツール性能はユーザー検証および反復的改善を通じて評価されました。モデル更新時には精度および有用性の改善が確認され、継続的改善アプローチが採用されています。	A	A	A	N/A	N/A
<b>透明性</b>	検索結果には参照元文書が提示され、生成された要約の根拠をユーザーが確認可能としています。また、プロンプト設計および利用ガイドラインが整備されています。	A	A	A	N/A	N/A

原則	活動内容	SPEC	DEV	PreD	PstD	RU
<b>データプライバシー</b>	プライベート環境およびベクトルストアを使用し、企業内データへのアクセス制御を実施しました。データ保護およびセキュリティ対策が導入されています。	A	A	A	N/A	N/A
<b>公正性と 公平性</b>	本ツールは文書検索および要約支援を目的とし、直接的な意思決定を行わないため、公平性リスクは限定的と評価されています。	A	A	A	N/A	N/A
<b>ガバナンスと アカウンタビ リティ</b>	開発段階ではプロジェクトチームが責任を負い、本番展開時にはPV組織およびQMS内での統合ガバナンスが想定されています。責任分担と管理体制は段階的に拡張される計画です。	A	A	A	N/A	N/A

## 略語

SPEC (Collection of specifications)：仕様および要求事項の収集

DEV (Development and change management)：開発および変更管理

PreD (Pre-deployment & post-change sign-off)：導入前および変更後承認

PstD (Post-deployment & post-change hyper-care)：導入後および変更後ハイパーケア

RU(Routine Use)：通常運用

A (Applicable)：適用あり

NA (Not Applicable)：該当なし

# ユースケース H: (ヒドロキシ) クロロキン網膜症の診断と予測を支援する人工知能

出典：11,12,13,14,15

PV領域：PV in The Clinic

## H1. ビジネス上の背景および課題

臨床における医薬品安全性監視（PV）は、個人における副作用の予防および治療に関係します。予防には一次予防が含まれ、これは副作用を予測し得る患者特性の潜在的に複雑または非明示的な組合せを特定することにより達成され、最適な薬剤選択（すなわち精密医療）を導くことを目的とします。また、二次予防および三次予防も含まれ、副作用の早期診断およびその後の介入を通じて、副作用の影響を軽減することを目的とします。以下に例を示します。

クロロキンおよびヒドロキシクロロキンはリウマチ学領域において重要な薬剤です。これらは他の治療選択肢と比較して比較的忍容性が高いものの、網膜毒性のリスクが存在し、早期に検出されない場合には重篤な視覚障害を引き起こす可能性があります。そのため、適時に薬剤を中止する必要があります。しかしながら、網膜症が診断された時点では不可逆的な網膜損傷がすでに発生している可能性があります。逆に、予測AIが進行に関する十分な先行指標を提供できる場合、治療期間およびそれに伴う治療上の利益を最大化できる可能性があります。歴史的に、スクリーニングおよび検出のゴールドスタンダードは眼底写真および自動視野検査でした。近年では、多局所網膜電図（mfERG）および光干渉断層計（OCT）が診断手段として追加されています。これらは通常、人間の読影者、理想的には網膜専門医によって評価されますが、時間的変化を含む微細な変化は見逃される可能性があります。また、すべての地域に必要な装置や網膜専門医が存在するわけではありません。不可逆的障害の発生前に網膜症を示唆する早期機能変化を特定、または進行をより良く予測するために、人間の視覚評価を補充することが理想的です。AIは、加齢黄斑変性（AMD）および糖尿病網膜症（DR）など、網膜画像に基づく複数の眼疾患の検出または予測において潜在的有用性を示しています。同様に、（ヒドロキシ）クロロキン網膜症の診断または予測のためのAIも、後ろ向き研究として開発および検証されています。

## H2. ソリューション

ヒドロキシクロロキン網膜症の診断のために、AIはカラー眼底写真、光干渉断層計（OCT）、および多局所網膜電図（mfERG）波形に適用されてきました。Fanらは、紹介専門クリニックにおける網膜症陽性患者（25例）および網膜症陰性患者（66例）から得られた176枚の眼底写真に対し、網膜症検出を目的として4種類のディープラーニングモデルを用いたハイパースペクトルイメージング（HIS）の研究を実施しました。Kulyabinらは、mfERGの生データに基づくディープラーニング分類と、従来のmfERG読出しパラメータに基づくモデルとを比較し、分類および視野感度の予測（回帰）を評価しました。この研究では、0.7～20.9年間にわたりmfERGおよび視野検査でモニタリングされた主として女性患者53例（網膜症陰性35例、最小網膜症9例、顕性網膜症9例）が対象となりました。Kalraらは、主として女性で構成される368名/388眼から取得した臨床特徴およびスペクトラルドメインOCT（SD-OCT）特徴を用い、ランダムフォレストを使用した自動診断および疾患進行予測を実施しました。Habibらは、1463適格眼（主として女性患者748名）に対してmfERGデータを用い、ヒドロキシクロロキン網膜症を同

定するためのサポートベクターマシン (SVM) を学習させました。このうち95眼 (48名) が対照群として組み入れられました。ごく最近では、Woodward-court らが、SD-OCT画像からヒドロキシクロロキン網膜症の存在検出および将来的発症予測を目的として、網膜症尤度スコア (Likelihood of Retinopathy Score : LRS) を算出する畳み込みニューラルネットワーク (CNN) の開発および適用を報告しました。この研究は、ヒドロキシクロロキン網膜症陽性171例および陰性238例を含む409名の患者、ならびに5つの独立した国際的な臨床施設から収集された8251件のSD-OCT Bスキャン (1988ボリューム) という、より大規模かつ多様なデータセットを使用している点で特筆されます。これらの施設は、自己申告による人種・民族的背景の多様性を反映しており、さらに2種類の異なるSD-OCT技術が含まれていました。

### H3. 結果

Fan らの研究において最も高い性能を示したディープラーニングモデルは、正確度、適合率、再現率、特異度、およびF1スコアがすべて0.95以上を達成しました。また、元の網膜画像と比較して、ハイバースペクトル画像を使用した場合により優れた性能が示されました。Habib らによるサポートベクターマシン (SVM) は、特異度84.0%、感度90.9%を示しました。性能は、スクリーニング用途では感度を重視し、診断用途では特異度を重視するよう調整可能でした。Kalra らは、検出において平均AUC 0.97、感度95%特異度91%を報告し、進行予測では平均AUC 0.89、再現率90%、特異度80%を報告しました。Kulyabin らは、完全なmfERG波形を使用したAIモデルにおいて、バランス精度最大0.795、適合率最大0.844、再現率最大0.866、F1スコア最大0.771を報告しました。Woodward-Court らは、CNNベースのアルゴリズムが臨床診断時点でヒドロキシクロロキン網膜症を検出できるだけでなく、臨床診断前の十分なリードタイムにおいても検出可能であることを報告しました (臨床診断前平均220.8日)。性能指標は、正確度0.987 [95%信頼区間：0.962–1.00]、感度1.00 [95%信頼区間：0.833–1.00]、特異度0.983 [95%信頼区間：0.952–1.00]、陽性的中率 (PPV) 0.944 [95%信頼区間：0.836–1.00]、陰性的中率1.00 [95%信頼区間：0.937–1.00] でした。網膜症を発症した眼において、臨床診断に対する平均リードタイムは2.74年でした。また、このアルゴリズムは、左右眼間におけるLRSの決定係数が高値 (0.93) であったこと、およびLRSの時間的推移が本網膜症の既知の臨床経過と一致していたことに基づき、表面的妥当性 (face validity) も示しました。

### H4. ガバナンスフレームワークへの適合

レビュー対象となった研究とガバナンス枠組みとの整合性を検討するにあたり、いくつかの点を最初に指摘します。これらの研究はいずれも後ろ向き研究であり、実現可能性評価またはパイロット研究として実施されたものであり、日常診療への応用に関する進展は報告されていません。

#### 課題および得られた教訓

Woodward-Court らによる最新の研究は、複数の点で既存研究に対する進展を示しましたが研究ごとに対象集団は比較的小規模であり、また限定的または不均衡な特性を有していました。薬剤使用の臨床状況が自己免疫疾患を含むことが多いため、被験者は主として女性でした。アジア人は研究サンプルにおいて過小代表であり、より大規模かつ多様な集団における追加的な評価の必要性が示されました。それでもなお、Woodward-Court らの研究は既存研究と比較してより大規模かつ多様であり、さらに2種類の異なるSD-OCT装置を用いた評価も含まれていました。地理的に異なる複数のデータセットを用いる場合、通常は装置差による変動が一般化可能性に対する課題となりますが、これが常

に十分に考慮されているとは限りません。網膜病変を対象とするAI診断研究でしばしば見られるように、網膜の併存疾患は除外または過小代表となっており、複数の網膜疾患（例：糖尿病網膜症および薬剤誘発性網膜症）を有するより多様な患者集団への一般化可能性を制限しています。また、観察単位として「眼」を使用していることは、擬似反復（pseudo-replication）の問題およびそれが性能推定に与える潜在的影響を生じさせる可能性があります。信頼区間は通常提示されていませんでした。

重要な点として、最も新しい研究では、医療システムの観点からの診断上の課題および望ましい特性の中で本ソリューションが強調されていました。すなわち、単一で広く利用可能かつ自動化可能な診断手段を用いることにより、患者および医療システムの負担を軽減し、さまざまな専門レベルの臨床医が診断を行えるようにする、いわば診断の「民主化」を目指すものです。

本疾患は稀であるため、この種の臨床研究には大規模な患者コホートが必要となり、前向き検証における大きな障壁となります。将来の臨床的網膜症発症予測を検証するためには、非常に長期間の患者追跡が必要です。さらに、多くの地域において医療機器ソフトウェアとしての規制承認取得には時間および費用を要するため、臨床現場への前向きの導入は依然として困難です。今後の研究としては、眼科診療におけるアルゴリズム導入の費用対効果評価を含め、将来的な開発判断を支援するための財務的評価が考えられます。

表19: ユースケースH: ガバナンスフレームワークとの整合性 (詳細)

出典：CIOMS Working Group XIV

原則	活動内容	SPEC	DEV	PreD	PstD	RU
リスクに基づくアプローチ	整合していない。これらのパイロット研究では、リスク評価およびリスク低減計画は提示されていませんでした。ただし、ヒューマン・イン・ザ・ループ (HITL) 枠組み内での配置については、1つ以上の研究において明示的に検討されていました。	N/A	N/A	N/A	N/A	N/A
ヒューマン・オーバーサイト	部分的整合。臨床環境における実現可能性／パイロット研究を報告した1つ以上の論文では、HITLなどヒューマン・オーバーサイトの観点から適切な導入について議論されています。しかし、変更管理および職員教育計画については議論されていません。また、一部地域では利用可能なヒューマン・オーバーサイトが網膜専門医より経験や専門性の低い一般臨床医によって提供される可能性があり、医療資源が限られた環境では段階的な利益をもたらす可能性がある点が述べられています。	A	A	N/A	N/A	N/A

原則	活動内容	SPEC	DEV	PreD	PstD	RU
<b>妥当性と頑健性</b>	<p>部分的整合。参照基準は定義されています。</p> <p>1つ以上の研究で、使用されたデータセットの不均衡が一般化可能性を損なう制限として指摘されています。また、いくつかの研究では他の眼疾患を有する患者が除外されており、分類は「ヒドロキシクロキン（HCQ）網膜症あり」対「正常網膜」の2群となっていました。このことは、網膜に影響を与える他の眼疾患を併存する患者に対するスクリーニングへの一般化可能性を制限します。</p> <p>さらに、対象母集団（導入対象領域）がすべての研究で明確に定義されていませんでした。データ前処理（例：網膜画像のトリミング）を日常運用へ統合することに関する議論も行われておらず、一部研究では観察単位が「眼」とされており、擬似反復（pseudo-replication）に関する疑問が生じています。</p>	A	A	A	N/A	N/A
<b>透明性</b>	<p>1つ以上の論文において、ヘルシンキ宣言の原則の遵守および倫理審査委員会（Institutional Review Board：IRB）の承認取得が報告されています。また、結果説明として特徴分布のヒートマップなどが提示された研究もあります。</p>	A	A	A	N/A	N/A
<b>データプライバシー</b>	<p>参照された1つ以上の研究において、ヘルシンキ宣言の原則の遵守および倫理審査委員会の承認取得が宣言されています。</p>	A	A	A	N/A	N/A
<b>公正性と公平性</b>	<p>参照された1つ以上の研究において、ヘルシンキ宣言の原則の遵守および倫理審査委員会承認の取得が報告されています。また、アジア人など特定集団がデータにおいて過小代表となっている可能性が指摘されており、より多様な集団を含む大規模データセットによる追加評価が推奨されています。</p> <p>さらに、網膜専門医が利用できない医療資源の限られた地域や代表性の低い地域における適用可能性についても議論されており、この点は前述のヒューマン・オーバーサイトの項目とも関連しています。</p>	A	A	A	N/A	N/A
<b>ガバナンスとアカウントビリティ</b>	<p>臨床環境で実施されたこれらの研究は、ヘルシンキ宣言の原則に従って実施され、それぞれの倫理審査委員会の承認を受けています。</p>	A	N/A	N/A	NA	N/A

## 略語

SPEC (Collection of specifications) : 仕様および要求事項の収集  
 DEV (Development and change management) : 開発および変更管理  
 PreD (Pre-deployment & post-change sign-off) : 導入前および変更後承認  
 PstD (Post-deployment & post-change hyper-care) : 導入後および変更後ハイパーケア  
 RU (Routine Use) : 通常運用  
 A (Applicable) : 適用あり  
 NA (Not Applicable) : 該当なし

## 付録 3 – 参考文献

- Roemming H-J, Hauben M, Wannhoff W, et al. How LLMs can advance safety case intake—points to consider and insights from a proof of concept. *Therapeutic Advances in Drug Safety*. 2025;16. doi:10.1177/20420986251386222 (Webpage accessed 12 December 2025)
- Kreimeyer K, Spiker J, Dang O, De S, et al. Deduplicating the FDA adverse event reporting system with a novel application of network-based grouping. *J Biomed Inform*. 2025;May;165:104824. doi:10.1016/j.jbi.2025.104824. (Journal full text)
- Kreimeyer K, Menschik D, Winiacki S, Paul W, et al. Using Probabilistic Record Linkage of Structured and Unstructured Data to Identify Duplicate Cases in Spontaneous Adverse Event Reporting Systems. *Drug Saf*. 2017;Jul;40(7):571-582. doi: 10.1007/s40264-017-0523-4. (Journal full text)
- Kreimeyer K, Dang O, Spiker J, Gish P, Weintraub J, Wu E, Ball R, Botsis T. Increased confidence in deduplication of drug safety reports with natural language processing of narratives at the US Food and Drug Administration. *Front Drug Saf Regul*. 2022;2:918897. doi:10.3389/fdsfr.2022.918897. (Journal full text)
- Römning H-J, Pushparajan R. AI Translation Assistant for Pharmacovigilance. Poster presented at DIA Europe 2021. (Full text accessed 21 March 2025)
- Papineni K, Roukos S, Ward T, Zhu WJ. BLEU: a method for automatic evaluation of machine translation. In: *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL); 2002 Jul; Philadelphia, PA*. Stroudsburg (PA): Association for Computational Linguistics; 2002. p. 311–318. (Webpage accessed 23 September 2024)
- Painter JL, Chalamalasetti VR, Kassekert R, Bate A. Automating pharmacovigilance evidence generation: using large language models to produce context-aware structured query language. *JAMIA open*. 2025;Feb;8(1):o0af003. <https://doi.org/10.1093/jamiaopen/o0af003> (Journal full text)
- Cherkas Y, Ide J, van Stekelenborg J. Leveraging Machine Learning to Facilitate Individual Case Causality Assessment of Adverse Drug Reactions. *Drug Saf*. 2022;45;571-582. <https://doi.org/10.1007/s40264-022-01163-6> (Journal abstract)
- Modified from Cherkas Y, Ide J, van Stekelenborg J. Leveraging Machine Learning to Facilitate Individual Case Causality Assessment of Adverse Drug Reactions. *Drug Saf*. 2022;45 571–582. <https://doi.org/10.1007/s40264-022-01163-6> (Journal abstract)
- Al-Azzawi F, Mahmoud I, Haguinet F, Bate A, Sessa M. Developing an Artificial Intelligence-Guided Signal Detection in the Food and Drug Administration Adverse Event Reporting System (FAERS): A Proof-of-Concept Study Using Galcanezumab and Simulated Data. *Drug Saf*. 2023;Aug;46(8):743-751. <https://doi.org/10.1007/s40264-023-01317-0> (Journal full text)
- Fan WS, Nguyen HT, Wang CY, Liang SW, Tsao YM, Lin FC, Wang HC. Detection of Hydroxychloroquine Retinopathy via Hyperspectral and Deep Learning through Ophthalmoscopy Images. *Diagnostics (Basel)*. 2023;Jul14;13(14):2373. <https://doi.org/10.3390/diagnostics13142373> (Journal full text)
- Kulyabin M, Kremers J, Holbach V, Maier A, Huchzermeyer C. Artificial intelligence for detection of retinal toxicity in chloroquine and hydroxychloroquine therapy using multifocal electroretinogram waveforms. *Sci Rep*. 2024;Oct22;14(1):24853. <https://doi.org/10.1038/s41598-024-76943-4> (Journal full text)
- Kalra G, Talcott KE, Kaiser S, Ugwuegbu O, Hu M, Srivastava SK, Ehlers JP. Machine learning–based automated detection of hydroxychloroquine toxicity and prediction of future toxicity using higher-order OCT biomarkers. *Ophthalmol Retina*. 2022;Dec1;6(12):1241-1252. <https://doi.org/10.1016/j.oret.2022.05.031> (Journal abstract)
- Habib F, Huang H, Gupta A, Wright T. MERCI: a machine learning approach to identifying hydroxychloroquine retinopathy using mfERG. *Doc Ophthalmol*. 2022;Aug;145(1):53-63. <https://doi.org/10.1007/s10633-022-09879-7>. (Journal abstract)
- Woodward-Court P, Hogg J, Lee T, Taribagil P, Zhao CS, et al. Deep learning algorithm for the diagnosis and prediction of hydroxychloroquine retinopathy: an international, multi-institutional study. *Ophthalmol Retina*. 2025;9(1):1-10. <https://doi.org/10.1016/j.oret.2025.06.003>. (Journal full text)

## 付録 4.

# 説明可能性ならびに公正性と衡平性に関する コンテンツ

## 説明可能性に関する具体的事例

第 6 章「透明性」に記載されているとおり、さまざまな医薬品安全性監視（PV）業務においてAIが「なぜ」「いつ」「どのように」使用されているかを開示することは不可欠です。これは、開発者、PV専門家および意思決定者、規制当局、医療従事者（HCP）、ならびに患者を含む関係者間における信頼、認識、および責任を維持するためです。しかしながら、説明可能性に対する要求事項およびその適用方法は状況に応じて異なります。例えば、誰が説明を求めているのか、その目的は何か、業務の性質、またはシステムライフサイクルのどの段階であるかによって異なります。<sup>1</sup>本節では、想定されるさまざまな状況を示すための具体例を提示します。これらの例で示された説明可能性の利点は要約され、本節の最後で統合的に整理されます。

### AI支援型医薬品安全性監視（PV）業務における説明可能性の例

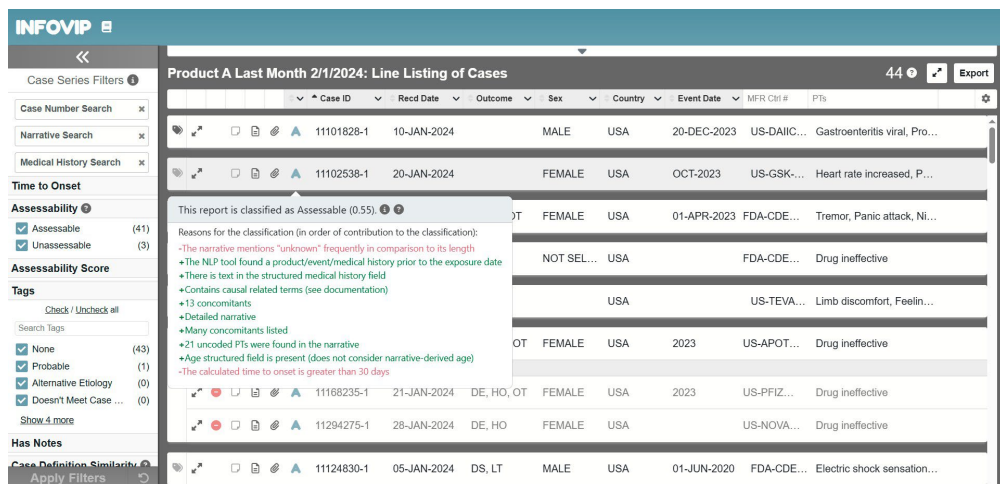
AIシステムによって注目対象として選択された症例をPV担当者がレビューしている状況を考えます。このとき、その分類理由が直ちには明確でない場合があります。このような状況では、症例データ中のどのテキストがAIの推奨判断に寄与したのかを示す情報へアクセスできることが、レビュー担当者にとって有益となります。実際の例を以下に示します。

米国FDA医薬品評価研究センター（CDER）のために開発された情報可視化プラットフォーム（Information Visualization Platform：InfoViP）は、AIシステムによって支援されるシグナル検出および評価において、説明可能性が人間の専門家にもたらす利点を示す例です。<sup>2</sup> InfoViPは自然言語処理（NLP）および複数の構成要素を用いて、市販後データを複数の情報源（FAERS、添付文書、生物医学文献）から処理し、情報の可視化、すなわち説明情報を提供します。これにより、米国FDAのFAERSデータベースに提出された数百万件の有害事象報告から潜在的シグナルを検出および評価する医療レビュアーを支援します。NLPコンポーネントである電子医療記録イベントベーステキストマイニング（Event-based Text-mining of Health Electronic Records：ETHER）は、最新のフロントエンド技術と組み合わせることで、症例ナラティブ内の関連テキストを色分け表示し、レビュアーがシグナル関連情報へ集中できるよう視覚的情報を提供します。

さらに、情報量が十分でレビュー支援に有用な症例を識別するモデルが、報告品質評価を支援するとともに、選択理由に関する具体的な説明を提供します。これらの機能は、症例重複排除および複数のフィルタリング機能と組み合わせられることで、数百万件の報告に対して人間のみでは実施困難な迅速レビューを可能にしています。<sup>3</sup>

図 8: 米国FDAの情報可視化プラットフォームのユーザーインターフェイス: アクセシビリティの分類に積極的に貢献する機能に焦点を当てシステムの機能を表示

出典: Botsis T et al, 2024<sup>3</sup>



上記の例は、Albahri ら (2023) によって述べられている説明可能性の中核的利点を示しています。<sup>4</sup> 説明可能性は、人間の専門家が「妥当で信頼できる」意思決定を行うことを支援します。最終的に、人間による判断およびそれに付随する説明が保持されることでシステム所有者やコンプライアンス確保を担うQA担当者の信頼を育むとともに、特定の症例がシグナルとして選択または除外された理由を確認しようとする規制当局からの信頼にもつながります。

さらに、説明情報により、誤った予測を生じさせているバイアスや疑似相関にユーザーが気付く可能性もあります。この情報を開発チームへ報告することで、将来的な改善に寄与します。このように、説明可能性は、導入後に発生し得るバイアスリスクや性能問題に対する継続的監視に有用であり、意思決定の信頼性を継続的に確保する役割を果たします。その結果、事後的説明可能性 (post hoc explainability) は、AI支援型意思決定に対する信頼および公平性の認識向上につながることを示されています。<sup>4</sup>

## 説明可能性から恩恵を受ける医薬品安全性監視 (PV) 関係者の例

PVの場面において一般市民が説明可能性を必要とする可能性は高くないかもしれませんが、AI利用が一般化するにつれて、その可能性を完全に排除することはできません。以下に想定されるいくつかの例を示します。

- 医療従事者 (HCP) または患者である報告者が重篤な有害事象を報告したにもかかわらず、AIトリアージシステムによって非重篤症例として処理された場合、報告者がMAHに対して説明を求める可能性があります。従来は、最終的なトリアージ判断を行ったPV担当者から説明を受けることが可能でした。しかし、この判断が自動化されたAIトリアージプロセスで行われた場合、説明可能性が欠如していると、報告者による結果への信頼および受容に影響を及ぼす可能性があります。

- 薬物相互作用の予防を目的として、薬剤師による薬物療法管理を支援するために GenAI が使用される場合、薬剤師および患者の双方が AI の推奨内容に直接接することになります。<sup>5</sup> この場合、AI の推奨内容に関して、両者から疑問や説明要求が提起される可能性があります。

## システム開発における説明可能性の例

AI システムを学習させるデータサイエンティストまたは機械学習エンジニアは、AI が特定の予測に到達する際にどの特徴量が使用されたか、あるいは学習データに存在するバイアスが明らかになる場合に、説明可能性から恩恵を受けます。特に、固有の説明可能性を欠く複雑なシステムでは、補助ツールが説明を提供することで、結果を「反転」させるために何を変更または除外すべきかを明らかにし、トラブルシューティングを支援できます。<sup>6</sup> しかし、多くの場合、このような知見に基づいてディープニューラルネットワークのシステム構造や特定の特徴量を調整することは容易ではありません。これらの説明可能性手法は、むしろ学習データに潜在する隠れたバイアスを特定し、それを修正する用途において有効である可能性が高く、以下の例で示されます。

Ribeiro ら (2016)<sup>7</sup> は、Local Interpretable Model-Agnostic Explanations (LIME) を活用して説明可能性を支援し、誤った予測の原因を明らかにできることを示しました。この実験では、犬とオオカミの画像を識別するモデルに対し、最初に意図的にオオカミを雪景色と関連付けて学習させました。すなわち、他の季節におけるオオカミ画像を除外することで、学習データに意図的なバイアスを導入しました。その結果、緑の背景にいるオオカミが犬と判定され、雪景色にいるハスキーがオオカミと判定されるという予測が生じました。LIME は、AI が予測時に特徴量として使用した画像領域を被験者に示し、誤認識の原因を特定できるかを検証しました。被験者は背景の雪が誤予測の原因となった特徴であることを正しく特定しました。この結果は、事後的説明可能性手法がブラックボックス的なディープニューラルネットワークによる予測を説明し、学習データに存在する問題および誤った相関関係を明らかにできることを示しています。

PV においても同様の手法を用い、AI が関連特徴として認識したテキスト中の単語を可視化することが可能です。実際の未公開事例では、AI トリアージシステムが一部の重篤症例を誤分類していた際、PV 分野の専門家 (SME) は、重篤性予測において AI がどの語句に着目しているかを確認することで有益な知見を得ました。このケースでは、LIME 分析により薬剤名への過度な注目が明らかとなりました。見逃された重篤症例が一般用医薬品 (OTC 薬) に関連していたことと合わせて、AI が「OTC 薬は重篤事象を起こしにくい」という疑似的な前提を学習していたことが判明しました。説明可能性から得られたインサイトに基づき、開発者は別モデルへの切替を検討し、OTC 薬に関連する重篤症例の不足など学習データのバイアスを検証し、または特徴量設計を通じて薬剤名を意思決定に使用しないよう AI へ指示することで問題を解決できました。

したがって、説明可能性は、誤予測の原因となる隠れたバイアス、特徴量、および疑似相関を明らかにすることで、開発者が AI モデルを評価する際に十分な情報に基づく判断を行うことを支援します。また、同一の学習データで訓練されたモデル間の性能差の要因を明らかにし、モデル選択にも寄与します。このプロセスを透明性のある形で文書化することは、開発者のみならず、システム所有者、利用者、さらには規制当局に対する信頼醸成にも大きく寄与します。

## 医療従事者および患者と相互作用する人工知能システムの例

仮想的な例として、販売承認保有者（MAH）が提供するチャットボットを通じて、医療従事者（HCP）が製品固有の情報を問い合わせるケースが考えられます。このようなチャットボットは、医薬品情報の提供から、有害事象（AE）や品質欠陥報告の収集まで、複数の目的を持つ可能性があります。医療従事者が、チャットボットの回答が不十分である、すなわち重要な医学用語や有害事象を考慮していない、あるいは疑問のある情報を提示していると判断した場合、説明を求めてMAHへ直接連絡する可能性があります。

上記は仮想例ですが、現在実際に利用可能なチャットボットの例として、WHOウェブサイト上の Smart Artificial Intelligence Resource Assistant for Health (SARAH) が挙げられます。SARAHは健康トピックに関する助言を提供することを目的としたプロトタイプ型チャットボットであり、ランディングページ上で医療助言を提供するものではないことが明確に示されています。<sup>8</sup> 一方で、SARAHは24時間365日、8言語で利用可能であり、公衆に対する有用なサービスの例でもあります。他方、メディアでは不正確または誤った情報を提示した事例、あるいは一部の質問に回答できなかった事例が報告され、OECDのAIインシデント監視データベースにも取り上げられています。<sup>9</sup> これは、チャットボットが導入された場合、患者、医療従事者、あるいはメディアが提供情報に対して異議を唱える可能性があることを示しています。したがって、PV領域においてMAHが一般市民と直接対話するAIソリューションを導入する場合、説明可能性の欠如は重要な検討事項となり得ます。

さらに、医療分野において一般市民と直接相互作用するシステムには、特別な注意が必要です。医療従事者であれば医学的誤りに気付く可能性が高い一方で、多くの一般消費者や患者はそれを識別できない場合があります。医学的背景を持たない個人は、誤った医療情報を受け入れ、それに基づいて行動してしまうリスクがあります。この点を示す例として、ChatGPTが提供する医療助言に対する信頼性を検討した研究では、医学的背景を持たない人々ほど、低リスクの健康トピックにおいてチャットボットを信頼する傾向が示されています。<sup>10</sup> 医学知識を欠く場合、誤情報を見抜けないことによる被害リスクが高まります。さらに、診断タスクにおいてGenAIモデルと医師を比較した83研究の系統的レビューおよびメタ解析では、AIモデルの全体診断精度は約52%（95%CI：47.0–57.1%）であり、非専門医と同程度ではあるものの、専門医より有意に劣る結果が示されています。<sup>11</sup>

したがって、一般市民がAI出力に疑義を呈した際に説明を提供できないという問題に加え、一般市民と直接接するAIシステムについては、そのリスクを十分に評価し、適切な低減策を講じることがシステム所有者に求められます。これは責任所在（accountability）の問題にも関係しており、個人に損害が生じた場合に責任を負うのはチャットボットではない点にも留意する必要があります。

### 説明可能性が利用できない、または必要とされない例

説明可能性が必ずしも必要ではなく、また技術的に実現困難な状況を示す例として、まず現在広く一般化している公開機械翻訳ツールの利用が挙げられます。一般的に利用者は、AIシステムが特定の文章をどのような内部プロセスで翻訳したのかについて、詳細な説明を求めません。また、GxP規制環境における翻訳では、翻訳が人間によるものであっても機械によるものであっても、品質チェック（QC）が必須となります。さらに、この品質チェックは通常、原言語と翻訳言語の双方に習熟した担当者によって実施されます。

このような状況では、品質確認を行う人間は翻訳エラーを検出できるだけでなく、機械翻訳における誤りの根本原因も理解できる場合があります。根本原因が自明な翻訳問題の例として、フィンランド語のような性別を持たない言語から英語へ翻訳する際に生じる性別バイアスがあります。<sup>12</sup> バイリンガルのレビュー担当者は、この性別バイアスを容易に認識し、AIによって導入された理由を理解した上で修正することが可能です。このような環境では、見かけ上正しい出力に繰り返し接することでレビュー担当者の注意力が低下し、機械翻訳に過度に依存してしまう「オートメーションバイアス (automation bias)」のリスクについても考慮が必要となります (第3章「リスクに基づくアプローチ」、第4章「ヒューマン・オーバーサイト」、第7章「データプライバシー」、第8章「公正性と衡平性」の各章も参照)。

PV領域における別の例として、Meldauら (2024) が報告した症例ナラティブの自動匿名化 (de-identification) が挙げられます。<sup>13</sup> この事例では、LLMを用いたシステムが症例記述中の人物名やイニシャルを自動的に検出し、マスキング (削除) する目的で利用されました。ヒューマン・イン・ザ・ループ (Human-in-the-Loop : HITL) 環境において、担当者はLLMがなぜ特定の文字列を人物名と判断したのかを理解できない場合がありますが、実務上それは必須ではありません。重要なのは、その情報を削除すべきかどうかを判断できる十分な性能、特に偽陰性 (見逃し) が少ないことです。一方で、この方法における説明可能性の欠如は、公平性と衡平性 (Fairness & Equity) の評価に課題をもたらしました。引用論文では、テストセット中で唯一匿名化されなかったフルネームがインド系の名前であったことが報告されており、説明可能性が不足している場合、こうした潜在的バイアスの検出や評価が困難になる可能性が示されています。

### 説明可能性が人間のプロセスおよび意思決定を改善し得る例

最後の例では、本質的に説明可能性を備えたAIモデルが、人間の業務プロセスおよび意思決定をどのように改善し得るかが示されています。有害事象 (AE) 報告における重複症例検出のための統計的手法の評価において、ノルウェーから報告された2件の症例が、AIモデルによって疑わしい重複として特定され、データセット内の他のすべての症例ペアよりも高い順位で提示されました。<sup>14</sup> しかし、これら2件は既知の重複としてラベル付けされておらず、人間の評価者にとっても明確な重複には見えませんでした。発症日や年齢は近いものの一致しておらず、有害事象用語も完全一致ではなく、臨床的に類似しているにとどまっていました。

AIモデルの出力内容を詳細に確認したところ、この症例ペアが重複の可能性が高いと分類された主因は、6種類の同一有効成分が完全一致していたことであると判明しました。これらの薬剤は通常同時報告されることが稀でした。その後、当該症例は同一病院の異なる医師によって報告された同一事象であることが国の規制当局によって確認され、これまで未認識だった重複報告であることが明らかになりました。<sup>14</sup> この事例では、人間の評価者は「6種類の薬剤が一致する」という事実がどれほど統計的に稀で重要な証拠であるかを十分に認識できておらず、評価において適切な重み付けを行っていませんでした。このようなインサイトは、将来的に人間オペレーターによる重複症例評価プロセスの改善に活用可能です。

### 説明可能性を支援する手法の例

本文書作成時点で利用されている代表的な事後の説明可能性 (post-hoc explainability) 手法には、以下が含まれます。

- LIME (Local Interpretable Model-Agnostic Explanations)  
本章前半で紹介された Ribeiro ら (2016) の例を参照。<sup>7</sup>

- SHAP (Shapley Additive exPlanations)  
Imran ら (2024) は、シグナル検証を支援する教師あり機械学習モデルにおいて SHAPを用いた説明可能性の例を提示しています。<sup>15</sup>
- 信頼スコア (Trust Score)  
モデル出力に対する不確実性を示す指標。<sup>16</sup>
- 信頼度スコア (Confidence Score)  
不確実性が高いアウトプットが人間のレビューの対象となるように目印を付けるために通用は利用可能となっている指標です。<sup>1</sup>
- 可視化 (Visualization)  
AI予測に寄与したテキスト部分のハイライト表示や、入力画像中で重要と判断された領域をヒートマップとして表示するサリエンシーマップ (saliency map) など。

PV領域では画像解析は主流ではありませんが、現在の技術動向を包括的に示す目的でサリエンシーマップも例として挙げられています (Plass ら, 2023)。<sup>16</sup>

PVでは主にテキストデータが扱われるため、視覚的説明は通常、症例データ内でAI判断に寄与したテキストをハイライト表示する形式となります。この点については、前述の米国FDAの InfoViP システムが、説明可能性の有益性を示す代表例として参照されています。<sup>2,3</sup>

## 公正性と衡平性に関するコンテンツ

以下に示す例のすべてがPVに特化したものではありませんが、不十分なデータ、過小代表集団に起因するバイアス、および明示的バイアスが、特定集団に対する不公平な取扱い、十分にサービスが提供されていない集団、ならびに治療の不平等につながり得ることを示しています。

### AIソリューションの学習が不十分であった、またはデータセットが不十分であったことにより不公平なバイアスが導入され、衡平性の欠如を招いた例

米国では、処方オピオイドは電子データベースである処方薬モニタリングプログラム (Prescription Drug Monitoring Programs : PDMPs) を通じて追跡されています。PVに特化した例ではありませんが、Bamboo HealthのNarxCare®は、PDMPを活用してオピオイド過量投与の可能性を予測するためのオピオイドリスク指標を算出するAI搭載ツールの一例です。医療上の意思決定を支援することを意図したツールであるにもかかわらず、複雑な病態を有し医療利用頻度が高い患者が「高リスクスコア」により疼痛管理において差別的に扱われ、十分なケアが提供されない可能性があるという観察が示されています。<sup>17</sup> このスコアはPDMPで入手可能な限定的データに基づいて算出され、他の要因は一切考慮されません。スコアに影響する要因の一つは処方医数です。教育病院で複数の医療提供者から治療を受けている患者は「処方医が多すぎる」と見なされ、複数の処方を得る目的で複数の医師から治療を求めていると解釈され得ます。2021年4月のDrug and Alcohol Dependence誌の研究では、「一般的なデータ駆動型アルゴリズム」が、複数の専門医を受診することが多いがん患者の20%を、複数のオピオイド処方を得る目的で複数の医師を探している患者として誤分類したことが報告されています。著者らが指摘しているとおり、PDMPデータには診断情報およびその他の重要な患者文脈が欠如しており乱用と適切な臨床使用を区別する能力が制限されています。2021年10月にDrug and Alcohol Dependence誌に掲載された研究では、独立したバリデーション研究が実施され、NarxCareツールの偽陽性率は17.2%、偽陰性率は13.4%であったと報告されています。<sup>18</sup>

この例では、サブグループ要因（例：複雑な病態を有する患者、複数処方医につながり得る医療提供体制、長期のオピオイド使用を要する患者に関する文脈の欠如、診断情報の欠如）を考慮しない不十分なデータによりNarxCareツールにバイアスが導入され、その結果、不適切かつ誤解を招く高い患者オピオイドリスクスコアの予測につながる可能性があります。バイアスにより高スコアが付与されるサブグループの患者における公平性および衡平性への脅威は、高スコアが単独で考慮された場合に、適切な疼痛管理が受けられなくなる可能性がある点にあります。

PVにおいて公平性および衡平性に対するリスクは主として、AIソリューションによって十分にサービスが提供されていないサブ集団に対し負の影響を及ぼし得る、または差別的な害をもたらし得る明示的バイアスから生じます。NarxCareの例はPVに関連するものではありませんが、不十分なデータと文脈の欠如に起因する明示的バイアス、ならびに「医療利用頻度が高い患者」に結び付けられた負のステレオタイプが適用されたことによる暗黙的バイアスの双方を示しています。<sup>17</sup>

### 過小代表集団に起因して生じたバイアスの例

ブラジルでは、人工ニューラルネットワークを用いた皮膚病変分類における判定精度について、コーカサス人患者とブラジル人患者の結果が比較されました。皮膚病変は、基本的なCNN（畳み込みニューラルネットワーク）アーキテクチャを用いて分類されました。ニューラルネットワークの学習にはInternational Skin Imaging Collaboration (ISIC) データベースが使用され、約25,000枚の皮膚病変画像がCNNに適用されました。これらの画像には、メラノーマ、色素性母斑、基底細胞癌、日光角化症、良性角化症、皮膚線維腫、血管性病変、および扁平上皮癌が含まれていました。

ISIC患者データを用いた試験では、約90%に近い精度が得られました。しかし、ブラジル人患者を対象とした試験では、皮膚病変検出の精度は40%未満にとどまり、コーカサス人患者における約90%の精度と比較して著しく低い結果となりました。その結果、CNNの検出精度が低いことにより、ブラジル人患者において皮膚病変の早期介入・予防的治療における不平等が生じる可能性が高くなることが示唆されました。<sup>19</sup>

### 明示的な負のバイアスの例

付録 4「システム開発における説明可能性の例」において、重篤症例を誤分類したAIトリアージシステムの事例が示されています。このAIソリューションは、学習データにおいて重篤事象が十分に含まれていなかったため、対象となるOTC医薬品に関連する有害事象を非重篤として予測する傾向を誤って学習しました。これは明示的な負のバイアスの例とも考えられます。

不十分な学習データセットに加えて、対象OTC製品では重篤な有害事象が発生する可能性は低いという明示的な先入観が存在していました。医療機関へのアクセスや医療従事者（HCP）への受診機会が限られている集団ではOTC製品への依存度が高い場合がありこれらの集団はマイノリティ集団に属する可能性が高いと考えられます。そのため、OTC製品に関連する重篤報告が体系的に非重篤として誤分類される場合、安全性リスクの特定および評価に影響を及ぼし得ることから、公平性および衡平性に対する重大な脅威となり得ます。

## 付録 4 – 参考文献

- 1 Royal Society. Explainable AI. Roy Soc [Internet]. 2024. ([Website](#) accessed 11 August 2024)
- 2 Spiker J, Kreimeyer K, Dang O, Boxwell D, Chan V, Cheng C, Gish P, Lardieri A, Wu E, De S, Naidoo J. Information visualization platform for postmarket surveillance decision support. *Drug Saf.* 2020;Sep;43:905-915. <https://doi.org/10.1007/s40264-020-00945-0> (Journal abstract)
- 3 Botsis T, Dang O, Kreimeyer K, Spiker J, De S, Ball R. A Decision-Support Platform Powered by AI and Humans-in-the-Loop Boosts Efficiency and Assures Quality in FDA’s Pharmacovigilance. International Society of Pharmacovigilance, 23rd Annual Meeting 2024, Montreal, Canada. ([Webpage](#) accessed 21 March 2025)
- 4 Albahri AS, Duhaim AM, Fadhel MA, Alnoor A, Baqer NS, Alzubaidi L, Albahri OS, Alamoodi AH, Bai J, Salhi A, Santamaria J. A systematic review of trustworthy and explainable artificial intelligence in healthcare: Assessment of quality, bias risk, and data fusion. *Inf Fusion.* 2023;Aug1;96:156-191. <https://doi.org/10.1016/j.inffus.2023.03.008> (Journal full text)
- 5 Roosan D, Padua P, Khan R, Khan H, Verzosa C, Wu Y. Effectiveness of ChatGPT in clinical pharmacy and the Role of Artificial Intelligence in medication therapy management. *J Am Pharm* 2023;Dec 2;64(2);422-428. <https://doi.org/10.1016/j.japh.2023.11.023> (Journal full text)
- 6 Hauben M. Artificial intelligence in pharmacovigilance: Do we need explainability?. *Pharmacoepidemiology and Drug Safety.* 2022;Dec;31(12):1311-1316. <https://doi.org/10.1002/pds.5501> (Journal full text)
- 7 Ribeiro MT, Singh S, Guestrin C. “Why Should I Trust You?”: Explaining the Predictions of Any Classifier [Internet]. arXiv:1602.04938 2016. <https://doi.org/10.48550/arXiv.1602.04938> (Journal full text)
- 8 World Health Organization (WHO). Using AI to lead a healthier lifestyle. WHO [Internet]. 2024. ([Website](#) accessed 21 March 2025).
- 9 Organisation for Economic Co-operation and Development (OECD). OECD legal instruments. Recommendation of the Council on Artificial Intelligence. OECD [Internet]. 2019. ([Webpage](#) accessed 21 March 2025)
- 10 Nov O, Singh N, Mann D. Putting ChatGPT’s medical advice to the (Turing) test: survey study. *JMIR Medical Education.* 2023;Jul10;9:e46939. <https://doi:10.2196/46939> (Journal full text)
- 11 Takita H, Kabata D, Walston SL, et al. A systematic review and meta-analysis of diagnostic performance comparison between generative AI and physicians. *npj Digit Med.* 2025;8:175 <https://doi.org/10.1038/s41746-025-01543-z> (Journal full text)
- 12 Savoldi B, Gaido M, Bentivogli L, Negri M, Turchi M. Gender bias in machine translation. *Trans Assoc Comput Linguist.* 2021;Aug;18;9:845-874. [https://doi.org/10.1162/tacl\\_a\\_00401](https://doi.org/10.1162/tacl_a_00401) (Journal full text)
- 13 Meldau EL, Bista S, Melgarejo-González C, Niklas Norén G. WHO Uppsala Monitoring Centre. Automated De-identification of Case Narratives Using Deep Neural Networks for UK Yellow Card [Internet]. ([Full text](#) accessed 21 March 2025)
- 14 Norén GN, Orre R, Bate A, Edwards IR. Duplicate detection in adverse drug reaction surveillance. *Data Min Knowl Discov.* 2007;14:305-328. <https://doi.org/10.1007/s10618-006-0052-8> (Journal abstract)
- 15 Imran M, Bhatti A, King DM, Lerch M, Dietrich J, Doron G, Manlik K. Supervised machine learning-based decision support for signal validation classification. *Drug Saf.* 2022;May;45(5):583-596. <https://doi.org/10.1007/s40264-022-01159-2> (Journal full text)
- 16 Plass M, Kargl M, Kiehl TR, Regitnig P, Geißler C, Evans T, Zerbe N, Carvalho R, Holzinger A, Müller H. Explainability and causability in digital pathology. *J Pathol Clin Res.* 2023;Jul;9(4):251-260. <https://doi.org/10.1002/cjp2.322> (Journal full text)
- 17 Siegel Z. In a world of stigma and bias, can a computer algorithm really predict overdose risk? *Ann Emerg Med.* 2022;Jun;79(6):A16-A19 <https://doi.org/10.1016/j.annemergmed.2022.04.006> (Journal full text)
- 18 Cochran G, Brown J, Yu Z, Frede S, Bryan MA, Ferguson A, et al. Validation and threshold identification of a prescription drug monitoring program clinical opioid risk metric with the WHO alcohol, smoking, and substance involvement screening test. *Drug Alcohol Depend.* 2021;228:109067 <https://doi:10.1016/j.drugalcdep.2021.109067> (Journal full text)
- 19 Artificial Neural Network Protocol in Dermatology. I-Workshop on Artificial Intelligence Applied to Health 2021 (I-WAIAH), 6th and 7th October 2021. (Feedback Report). ([PDF](#) accessed 29 September 2025)

## 付録 5.

## CIOMS 作業部会のメンバーと会合

安全性監視における人工知能に関するCIOMSワーキンググループXIVには、アカデミア、製薬会社、規制当局、国内および国際機関のステークホルダー・グループが参加しました。報告書の作成プロセスを文書化した議事録は、CIOMSのウェブサイト [www.cioms.ch](http://www.cioms.ch) で見ることができます。

学界		
氏名	会社/組織	国
Altman, Russ	Stanford University	米国
Botsis, Taxiarchis	Johns Hopkins University School of Medicine	米国
Dogné, Jean-Michel	University of Namur	ベルギー

製薬会社		
氏名	会社/組織	国
Amelio, Justyna	AbbVie	英国
Barrios, Luisa	Merck Sharp & Dohme	コロンビア
Bate, Andrew	GSK	英国
Bellur, Arvind	CSL Behring	米国
Berridge, Adrian	Takeda Development Center Americas, Inc	米国
Carroll, Hua	Biogen	米国
Cherkas, Yauheniya	Johnson & Johnson	米国
Cooper, Selin	AbbVie	英国
Diniz, Mariane	Bayer	ブラジル
Domalik, Douglas	AstraZeneca	英国
Franco, Piero Francesco	Pfizer	イタリア
Girod, Julie	Sanofi	米国
Grabowski, Neal	Sanofi	米国
Hauben, Manfred	Merck KGaA, Darmstadt, Germany	米国
Henn, Thomas	United Therapeutics	米国
Kara, Vijay	GSK	英国
Kempf, Dieter	Genentech	米国
Kidos, Kostadinos	Formerly Takeda Development Center Americas, Inc	米国

製薬会社		
氏名	会社/組織	国
Lorenz, Denny	Formerly Bayer AG	ドイツ
MacEntee Pileggi, Elizabeth	Johnson & Johnson	米国
Patel, Ravi	United Therapeutics	米国
Reinhard Pietzsch, John	Bayer	ドイツ
Römming, Hans-Jörg	Merck KGaA, Darmstadt, Germany	ドイツ
Straus, Walter	Moderna	米国
Whitehead, James	AstraZeneca	英国

規制当局		
氏名	会社/組織	国
Buch, Brian	Medicines and Healthcare products Regulatory Agency (MHRA)	英国
Durand, Julie	European Medicines Agency (EMA)	オランダ
Egebjerg Juul, Kirsten	Danish Medicines Agency (DKMA)	デンマーク
Harrison, Kendal	Medicines and Healthcare products Regulatory Agency (MHRA)	英国
Hirokawa-Voorburg, Satoko	Health and Youth Care Inspectorate (HYCI)	オランダ
Horst, Alexander	Swissmedic	スイス
Jensen, Morten	Danish Medicines Agency (DKMA)	デンマーク
Kjær, Jesper	Formerly Danish Medicines Agency (DKMA)	デンマーク
Ling, Benny	Health Canada	カナダ
Da Luz Carvalho Soares, Monica	Brazilian Health Regulatory Agency (ANVISA)	ブラジル
Matsunaga, Yusuke	Pharmaceuticals and Medical Devices Agency (PMDA)	日本
Mentzer, Dirk	Paul-Ehrlich-Institut (PEI)	ドイツ
Messelhäußer, Manuela	Formerly Paul-Ehrlich-Institut (PEI)	ドイツ
Moreira Cruz, Flávia	Brazilian Health Regulatory Agency (ANVISA)	ブラジル
Perez, Nicolas	Swissmedic	スイス
Scholz, Irene	Swissmedic	スイス
Stammschulte, Thomas	Swissmedic	スイス

規制当局		
氏名	会社/組織	国
Tregunno, Phil	Medicines and Healthcare products Regulatory Agency (MHRA)	英国

国内および国際的な組織・団体		
氏名	会社/組織	国
Mathur, Roli	Indian Council of Medical Research	インド
Meldau, Eva-Lisa	Uppsala Monitoring Centre/World Health Organization	スウェーデン
Norén, Niklas	Uppsala Monitoring Centre/World Health Organization	スウェーデン
Rosenfeld, Stephen	North Star Review Board	米国
Yau, Brian	World Health Organization	スイス

CIOMS		
氏名	会社/組織	国
Heaton, Stephen	Individual expert	ドイツ
Hill, Sanna	CIOMS	スイス
Le Louët, Hervé	CIOMS	スイス
Rägo, Lembit	CIOMS	スイス
Rannula, Kateriina	CIOMS	エストニア
Tsintis, Panos	CIOMS	英国

第XIV作業部会は、2022年から2025年までの間に以下のとおり11回開催されました。

第1章：スイス・ジュネーブ	2022年5月18-19日
第2章：スイス・ジュネーブ	2022年10月10-11日
第3章：バーチャルミーティング	2023年1月19日
第4章：バーチャルミーティング	2023年4月12日
第5章：チューリッヒ	2023年6月6-7日
第6章：バーチャルミーティング	2023年11月8日
第7章：バーチャルミーティング	2023年1月11日
第8章：ジュネーブ	2024年3月7-8日
第9章：ダルムシュタット	2024年9月24日~25日
第10章：スイス、ジュネーブ	2025年6月25日~26日
第11章：バーチャル会議	2025年9月8日

CIOMS 作業部会XIV編集チームは、2024年3月から2025年10月まで28回会合しました。

## 付録 6.

# パブリック・コンサルテーションの コメントーター

氏名	会社/組織	国/地域
Adeleye, Qadri	National Postgraduate Medical College of Nigeria	ナイジェリア
Audibert, Francois	Vitrana Inc.	米国
Biswas, Aditya	Johnson and Johnson	米国
Bryant, Jason	ArisGlobal	英国
Burns, Ashley	PIC/S Good Pharmacovigilance Practices Expert Circle Working Group on AI and Machine Learning Deputy Chairperson, FDA	米国
Chaturvedi, Tanvi	Soterius, Inc.	インド
Cho, Sylvia	US FDA	米国
Chouiyakh, Maria	Mohammed V University	モロッコ
Dave, Jay	COD Research PVT LTD	米国
de la Peña Solís, Francisco José	Novo Nordisk	メキシコ
El Hussien, Amira	RAY CRO	エジプト
Ezzeldin, Hussein	US FDA	米国
Freixas, Elisabet	Bristol Myers Squibb	スイス
Ghimire, Namita	Nepal Health Research Council	ネパール
Grigolo, Sabrina	EUPATI	イタリア
Gutierrez, Israel	Caparna Inc.	米国
Hapani, Kalindi	COD Research PVT LTD	インド
Heitmann, Martin	The Triality Group, LLC	ドイツ
Ho, Jeffrey	Perigent	英国
Iyer, Anand	Johnson and Johnson	米国
Jakubczyk, Jan	PIC/S Good Pharmacovigilance Practices Working Group on AI and Machine Learning, Polish Chief Pharmaceutical Inspectorate	ポーランド
Josephson, Aaron	Teva Pharmaceuticals	米国
Kessi, Una	HDI Safety, Oracle Health and Life Sciences	英国
Klueglich, Matthias	DGPharMed	ドイツ
Layton, Debbie	Lane Clark & Peacock LLP	英国
McAteer, Kaitlyn	Merck Animal Health	米国

氏名	会社/組織	国/地域
Nedog, Katarina	European Federation of Pharmaceutical Industries and Associations	ベルギー
Nilaus Præstegaard, Søren	Novo Nordisk A/S	デンマーク
Nishizawa, Claudio	ANVISA	ブラジル
Orriss, Andrew	Kenvue	米国
Patel, Jiggar	Kenvue	米国
Pharmacovigilance Working Group	The European CRO Federation, EUCROF	欧州
Prajapati, Vatsal	COD Research PVT LTD	インド
Prendergast, Christine	PIC/S Good Pharmacovigilance Practices Working Group on AI and Machine Learning, HPRA	アイルランド
Radicke, Sophie	PIC/S Good Pharmacovigilance Practices Working Group on AI and Machine Learning, MHRA	英国
Sahu, Aneesha	US FDA	米国
Salem, Myriam	PIC/S Good Pharmacovigilance Practices Expert Circle Working Group on AI and Machine Learning Chairperson, Health Canada	カナダ
Santana-Quintero, Luis	US FDA	米国
Schaeffer, Brian	Johnson & Johnson	米国
Scheerlinck, Rudi	Merck KGaA	ドイツ
Shee, Angela	Johnson & Johnson	米国
Singh Bedi, Simranjeet	Accenture Solutions Private Limited	インド
Smith, Sean	US FDA	米国
Stockton, Brandi	The Triality Group, LLC	ドイツ
Thomas, Michael	American Society of Pharmacovigilance Physicians	米国
Trevett, Kiernan	Roche	米国
Tsvetanova, Antonia	Lane Clark & Peacock LLP	英国
van Hunsel, Florence	Pharmacovigilance Centre Lareb	オランダ
Viñas, Norbert	Vigintake SL	スペイン
WENG, Xinyu	PIC/S Good Pharmacovigilance Practices Working Group on AI and Machine Learning, WHO	スイス
Wilson, Marie-Claire	Bristol Myers Squibb	スイス
Yuen, Alexander	Bristol Myers Squibb	スイス

氏名	会社/組織	国/地域
Zdorovtsova, Natalia	Lane Clark & Peacock LLP	英国
Zhou, Jessica	US FDA	米国
Anonymous	Sanofi	米国
Anonymous	Jazz Pharmaceuticals	イタリア

本報告書『医薬品安全性監視における人工知能』は、医薬品安全性監視（PV）、コンピュータサイエンス、数学、規制、法学、医学、人権、心理学および社会科学が交差する、急速に発展しつつある学際的分野を扱っています。したがって、医薬品の場合と同様に、医薬品安全性監視における人工知能の利用についても、その承認された適応、用法・用量、副作用、ならびに使用上の警告および注意事項を確立することが重要です。これらは、異なる背景を持つ多くの人々によって明確に定義され、理解されなければならず、それによって研究と実装を効果的、安全かつ責任ある形で推進することが可能になります。対象には、バイオ医薬品業界、規制当局および学界において医薬品安全性監視に従事する専門家、研究者、意思決定者が含まれます。また、シグナル管理や個別症例安全性報告（ICSR）処理のあらゆる側面を含め、医薬品安全性監視のための人工知能ソリューションを開発するソフトウェアベンダーも含まれます。本報告書は、応用科学研究や公的な議論への参加、あるいは各組織における評価や意思決定を通じて、この領域に積極的に関与するために必要な用語と概念的提供を提供します。

医薬品安全性監視における人工知能。CIOMS Working Group XIV の報告書。ジュネーブ：国際医学団体協議会 (CIOMS) 2025年。

この出版物は CIOMS のウェブサイトですべて無料で入手できます。

CIOMS の出版物は <https://cioms.ch/publications/> の出版物モジュールから入手できます。CIOMS, P.O. Box 2100, CH1211 Geneva 2, Switzerland, [www.cioms.ch](http://www.cioms.ch), email: [info@cioms.ch](mailto:info@cioms.ch).

