

データマネジメントにおけるArtificial Intelligenceの活用

～ これから始めるAI ～

日本製薬工業協会

データサイエンス部会

2022年度タスクフォース 1-1 「データマネジメントにおけるAIの活用」

2023年5月

目次

はじめに	5
1. データマネジメント領域におけるユースケース.....	6
1.1. コーディング.....	6
1.2. データレビューならびにクエリ文作成①：Pfizer 社の事例	9
1.3. データレビューならびにクエリ文作成②.....	13
2. AI リテラシー向上・学習データの準備・バリデーションについて	15
2.1. AI リテラシー向上の重要性	15
2.2. 学習データの準備に関する課題.....	16
2.3. システムバリデーション.....	17
3. 規制・倫理・法律上の課題.....	17
3.1. AI 原則、AI に係る規制確立に向けた動き.....	18
3.2. AI に係る医療分野での規制の現状	18
3.3. 倫理上の留意事項.....	19
3.4. 法律上の留意事項.....	20
3.4.1. 概論.....	20
3.4.2. 著作権法、特許法.....	21
3.4.3. 個人情報保護法.....	22
3.5. AI 利活用に伴う説明責任について	23
4. AI 導入検討方針および投資	24
4.1. AI 導入検討の方針	24
4.2. AI に対する投資対効果の評価	25
5. AI に取り組むためのヒント	27
5.1. 概説.....	27
5.2. (参考) 数理・データサイエンス・AI 教育プログラム認定制度	27
5.3. 参考文献の紹介.....	28
5.4. 教育プログラムの紹介.....	28
5.4.1. UTokyo OpenCourseWare extra (UTokyo OCWx).....	28
5.4.2. Kaggle.....	29
5.4.3. Papers with code.....	29
5.4.4. Python 学習教材	29
5.4.5. Coursera (コーセラ)	30
6. データマネジメントにおける AI 活用の展望.....	30
6.1. 臨床開発への応用.....	30
6.2. 臨床試験における AI	31
6.2.1. eCOA 入力タスクに関する通知を最適化する AI.....	32

6.2.2.	画像データの適切性をタイムリーにフィードバックする AI	32
6.2.3.	少数のセンサーで詳細な測定を可能にする AI	32
6.3.	PMS (Post Marketing Surveillance) における AI	33
6.3.1.	PMS の進捗管理	33
6.3.2.	ヘルプデスク	33
6.3.3.	有害事象の検知	34
	おわりに	38
Appendix 1:	緒論	40
A1.1.	機械学習	40
A1.1.1.	教師あり学習の手法	40
A1.1.2.	モデルについて	44
A1.1.3.	モデルの構築と評価	44
A1.2.	自然言語処理	45
A1.2.1.	データの前処理	45
A1.2.2.	形態素解析	46
A1.2.3.	文章のベクトル化	47
A1.2.4.	BERT (Bidirectional Encoder Representations from Transformers)	47
A1.2.5.	固有表現抽出	48
Appendix 2:	自然言語処理技術を活用した有害事象判定モデルの作成	49
A2.1.	プログラムの作成と実行	49
A2.1.1.	データの用意	49
A2.1.2.	文章のベクトル化	50
A2.1.3.	判定モデルの構築	52
A2.1.4.	評価結果	55
A2.1.5.	考察	57
A2.2.	プロジェクトへの実装に向けて	59
	参考文献	60

本報告書における用語とその説明

用語	説明
AI	2019年5月に発行された「 AIってなに? 」と同じく、「データから機械的にルールを見つけ出して情報処理をするもの、またはその技術」と捉える。
学習データ (training data)	機械学習のモデル構築において、パラメーターを自動調整するために用いるデータまたはデータセット。
検証データ (validation data)	学習データを用いて構築された学習済みモデルの初期評価を行い、ハイパーパラメーターを調整するために用いるデータまたはデータセット。
テストデータ (test data)	学習済みモデルを評価するために用いるデータまたはデータセット。検証データと異なり、学習済みモデルへのフィードバックを目的としない。
ハイパーパラメーター	学習データを用いたモデルの構築前に、人があらかじめ決定する値。モデル構築の学習過程で自動調整されるパラメーターに対して、ハイパーパラメーターは学習を行う前に人が設定する。

はじめに

内閣府の「AI戦略2022」では、自然災害、パンデミック、人口減少と高齢化など、現在の日本が総力を挙げて対応すべきさまざまな課題に対して、「これまでの閉塞を破る起爆剤として大きく活用すべき」ものとしてAIを捉えている [1]。

医薬品の研究開発においても、AIは前臨床開発から臨床試験データの記録と分析、ファーマコビジランス、臨床使用の最適化に至るまで、医薬品のライフサイクルのすべての段階で適用が進められている。臨床開発における利用という観点では、膨大な時間とコストを要する治験プロセスを大幅に合理化できる手法としてAIに期待を寄せる製薬会社も少なくなく、実際の臨床試験でも既に利用が開始されている。

新型コロナウイルス感染症に対するワクチンの早期承認あるいは緊急使用許可が期待される中、Pfizer社はAIを利用したCRFデータの潜在的な不整合の検出とそれに応じたクエリ文の自動生成プロセスを導入し、それによってデータレビューサイクルタイムが大幅に短縮され、早期データベース固定の実現に大きなインパクトを与えたと述べている [2]。Pfizer社以外にもデータレビューおよびクエリ文の自動生成にAIを活用している事例が報告されている [3]。またWHOの国際医薬品モニタリングを支援しているUppsala Monitoring Centreではその業務上大量の処理を必要とする薬剤辞書のコーディングにAIを活用していることを報告している [4]。このように臨床開発のデータマネジメント分野においてもAIを活用した事例が出てきている。

しかし、この分野でAIの利用が十分には進んでいるとは言えず、概念的にはAIが活用できるようなプロセスでも従来のプロセスがそのまま採用され続けているということがあられると思われる。この原因としてはAIに関する教育不足、人材不足もあげられるだろう。特に我が国ではAIの利活用を支える関連人材が不十分となっていることが問題とされている [1] [5]。AI開発のためには線形代数学や確率統計学の知識や、高度なプログラミングスキルも必要なため、専門家を活用してAIを導入していくことになることも少なくないと思われる。しかし専門家に相談するにしても、一定程度の知識は必要なため、効率的・効果的な学習方法が重要となってくる。

多くの科学技術と同様に、AIは社会に多大なる便益をもたらす一方で、大きなネガティブな影響を社会に与えるリスクも認識されている。リテラシー、プライバシー、セキュリティ、公平性、説明責任、透明性等の問題もAIの利用を進めていく上では十分に留意していく必要がある [6]。またデータセットやプログラムを利用する場合には、個人情報保護法、特許法や著作権法を含む知的財産法等についても留意し、他者の権利の侵害防止に努めなければならない。

臨床開発におけるAIの利用について、現時点で明確な規制要件は日欧米を含む規制当局から出されていない。しかし、アルゴリズムとその透明性確保が困難であることやAIの不具合のリスクが医薬品開発におけるAI利用ならびに患者の健康に影響を与える可能性などが規制上の課題として挙げられている [7]。そのため今後の規制動向をも注視していく必要

がある。

さて、2019年5月にデータサイエンス部会から発行された「AIってなに？」は、AIの基本から将来的な利用の展望まで包括的に扱い、またそれに続き同部会から2020年5月に発行された「機械学習始めました」は、機械学習プロジェクトの進め方を留意事項や実装例も含めて丁寧に説明しており、共に多くの読者により参照された有用性の高い書物であった。しかし、AIをデータマネジメント領域の実際の業務で活用していくためには、具体的な実装をイメージできるような橋渡しのガイドが必要と思われた。そこで本成果物では1章として具体的なイメージが掴めるようにデータマネジメント領域での実装例を紹介することにした。また既にこの「はじめに」で述べたような背景も踏まえ、データマネジメント領域に特化したことではないが、2章でAIリテラシー向上・学習データの準備・バリデーション、3章で規制、倫理、法律上の課題、4章でAI導入の方針と投資、5章でAIについて学ぶためのヒントについて概説することにした。6章ではデータマネジメント領域での今後の利用の展望に関する情報や考察を記載することにした。

本文に加えて二つのAppendixを用意してある。Appendix 1では「緒論」として、AIの技術用語について「AIってなに？」の記載内容をおさらいするとともに臨床試験への応用を概説することにした。また、Appendix 2では本タスクフォースでプロトタイピングとして実施した「自然言語処理技術を活用した有害事象判定モデルの作成」の試みを紹介することにした。データマネジメントに携わる全員が、ここで紹介した事例を自分自身で実現できるスキルを身に付ける必要はないと思われるが、このような知識はAI時代には必要になってくると思われることから、少なくとも必要なプロセス等は十分理解できるように学習されることを推奨する。

本書が臨床開発から製造販売後調査に関わる方々、特にデータマネジメント領域の業務に携わる方々にとってAIの利用促進のための参考となれば幸いである。

なお、本書で引用等行っているURLの最終参照日は全て2023年5月15日である。

1. データマネジメント領域におけるユースケース

1.1. コーディング

ここで述べるコーディングとは、医師がCRFにて報告した用語を辞書で用意された標準的な用語に結び付けることを指している。コーディングは2名の担当者により実施されることが多く、多くのリソースやコストがかかり、また担当者による偏りなどの問題も生じている。AIの利用によりその課題を解決するというアイデアは多くのデータマネージャーが考えてきたことと思われる。

Uppsala Monitoring Center (UMC) では、このような問題を解決するために2019年よりWHODrug KodaというAIを活用したシステムを薬剤情報のコーディングに利用している。本章ではシステムとしてのWHODrug Kodaの機能、外観、利用方法などの詳細で

はなく、データマネジメント領域で既に利用されている AI のユースケースとしてその基本的な仕組みを中心に紹介を行う [8] [9] [10] [11]。

WHODrug Koda は「報告された薬剤名」(必須)、「投与経路」、「適応」から薬剤名のコーディングを行い、また ATC コードを選択するシステムである (図 1-1-1)。

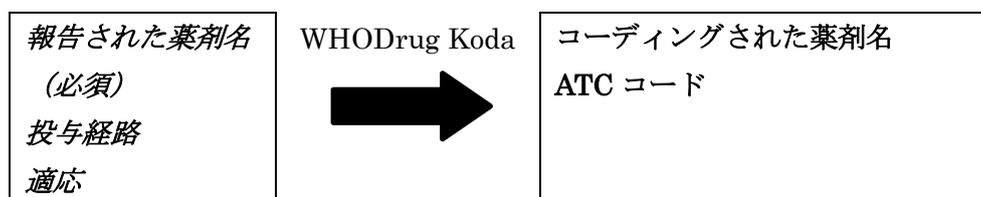


図 1-1-1

Coding Path (ダイレクトヒットや、同名の医薬品があったが適応によりコーディングを行った等) を提示して、High Certainty/Low Certainty

として結果を表示することにより、WHODrug Koda の結果に対する信頼度をユーザーにフィードバックする仕組みをもっている。

WHODrug Koda は報告された薬剤名の中にある剤型や濃度のような情報を特定し、それらを除いて報告された薬剤名から WHODrug の薬剤名を予測する (図 1-1-2)。

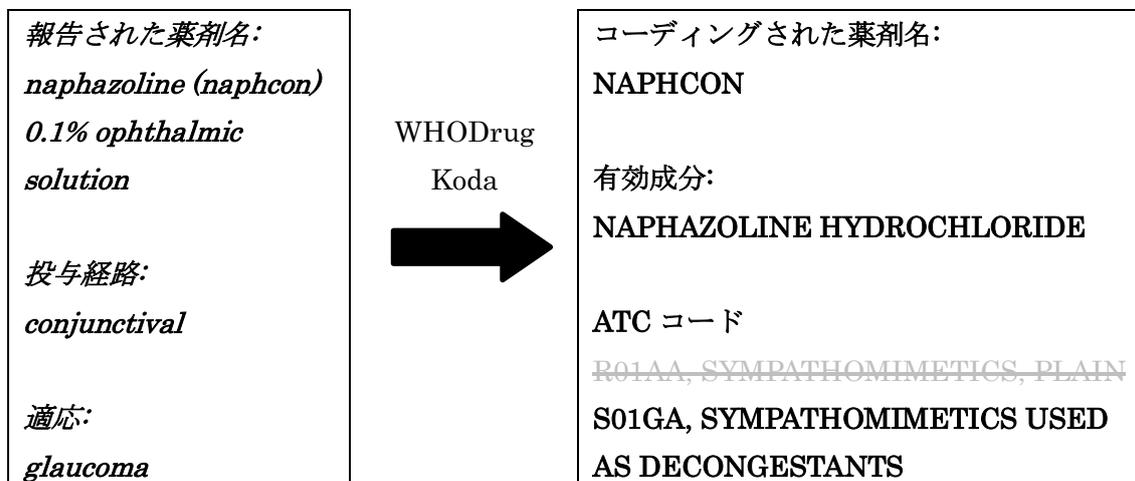


図 1-1-2

WHODrug Koda には教師あり機械学習が使われていて、それは薬剤名の選択にも用いられているが、主には ATC コードの選択に使用されている。機械学習には質のよい学習データが必要だが、WHODrug Koda の学習には UMC 自身がホストしている個別症例安全性報告の WHO データベースが用いられた。このデータベース中にある表 1-1-1 の例のようなコーディング済みの事例が数百万も存在しており、このことが WHODrug Koda 開発の重要な成功要因となった。

表 1-1-1

処理前のデータ			WHO データベースのコーディング	
報告用語	経路	適応	薬剤名	ATC コード
Acetazolamide	Oral	Cataract surgery	ACETAZOLAMIDE	S01EC, CARBONIC ANHYDRASE INHIBITORS
Acetazolamide Tablet 250mg	Unknown	Retinitis central	ACETAZOLAMIDE	S01EC, CARBONIC ANHYDRASE INHIBITORS
Acetazolamide	Orally	Secondary hypertension	ACETAZOLAMIDE	C03, DIURETICS
Acetazolamide	Oral	Seizure	ACETAZOLAMIDE	N03AX, OTHER ANTIEPILEPTICS

UMC では日々 WHODrug のメンテナンスを行うとともに、新しい安全性報告データのためのコーディングを行っており、それが WHODrug Koda の再教育に役立った。

WHODrug Koda の自身の結果に対する信頼度は、入力するデータがどのようなものかによるため絶対的な結果でないが、数多くのテストを試みた結果多少異なるものの、薬剤名のコーディングについては、High Certainty でコーディングした率が約 80%であり、Low Certainty でコーディングした率が約 15%であり、残りの約 5%は Uncoded であった (図 1-1-3)。ATC コーディングについては、High Certainty でコーディングした率が約 80%であり、Low Certainty でコーディングした率は約 20%であった (図 1-1-4)。また、WHODrug Koda は確かに間違えることがあるが、人間のコーディングチームにより実施され確認されたコーディング結果と比較した場合、その精度は High Certainty の結果では 99%を超えていて、人間のコーディング結果の適合率 (94-97%) と同等以上の精度が得られている。

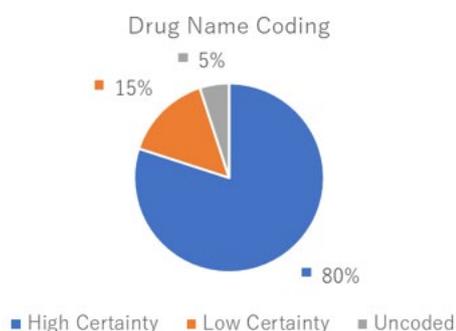


図 1-1-3

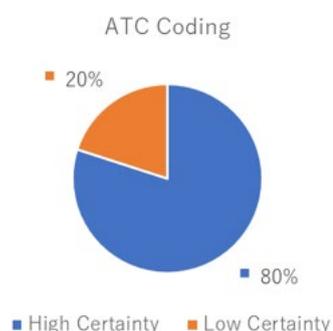


図 1-1-4

2022 年 9 月のアップデートにより、従来は Uncoded になっていた以下のようなミス

ペルを含む用語のコーディングも Low Certainty としてコーディング候補を出力できるようになっている（表 1-1-2）。

表 1-1-2

報告用語	薬剤名	確度
LACIDIPIN	LACIDIPINA	Low Certainty
BIKALUTAMID	BICALUTAMID	Low Certainty
OMEPRAZOLOM	OMEPRAZOLO	Low Certainty
LERCANIDIPIN	LERCANIDIPINA [LERCANIDIPINE HYDROCHLORIDE]	Low Certainty
	LERCANIDIPINA [LERCANIDIPINE]	Low Certainty
	LERCANIDIPINE	Low Certainty
MIZOLASTIN	MIZOLASTINA	Low Certainty
GLIMPERIDE	GLIPERIDE	Low Certainty

UMC の計算によると、人間のみが行う場合 1 万の薬剤名に対するコーディングおよび ATC コード選択に 558 時間を要するが、WHODrug Koda を用いた場合その所要時間が 153 時間に減少するということである。WHODrug Koda のコーディング結果の確度が Low Certainty や Uncoded の場合には人間が確認をする必要があるためそのための時間は要するが、コーディングの効率化という目的は WHODrug Koda 導入により一定程度達成できたものと思われる。

WHODrug Koda の情報は UMC のホームページ (<https://who-umc.org/>) の「Our products」にて参照できるので興味のある方は参照されたい

1.2. データレビューならびにクエリ文作成①：Pfizer社の事例

本タスクフォースチームのメンバーの所属会社である Pfizer 社では、収集データの中から齟齬の可能性のある個所を検出し、それに対応したクエリ文を作成するという一連のプロセスに Smart Data Query (SDQ) という AI を利用した仕組みを導入しているため、そのユースケースを紹介する [12] [13]。なお、本章の内容は Pfizer 社の公式見解を示すものではないことを予めお断りする。

SDQ の概略を図 1-2-1 に示す（なお、図中の Non-CRF は臨床検査や eCOA 等外部データを示している）。

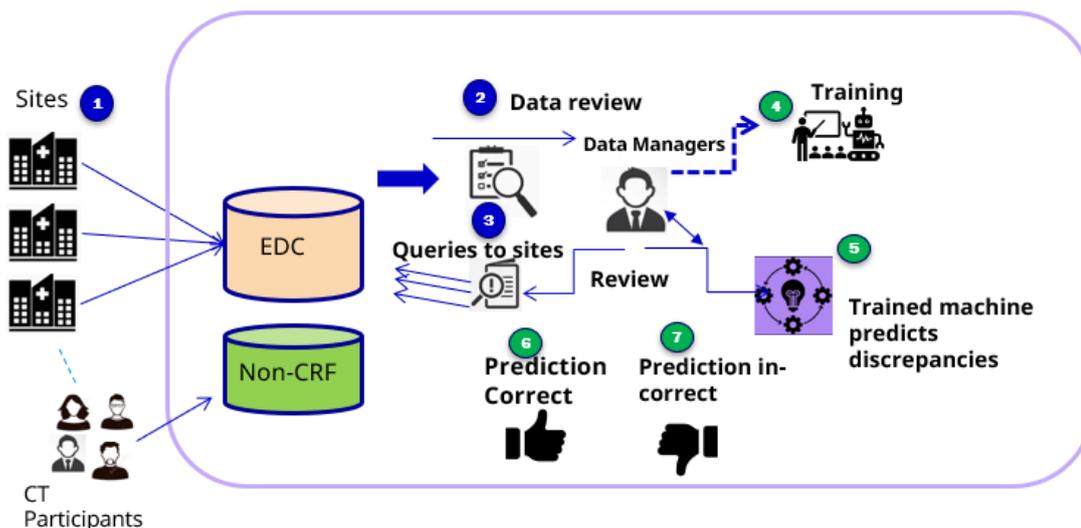


図 1-2-1

医療機関により EDC に入力されたデータや臨床検査や eCOA のような外部データに対して、AI がデータレビューを実施、齟齬を予測した上でそれに対応するクエリ文を生成する。次いで AI が予測した結果を人間が判断して、真の齟齬である場合に医療機関に問い合わせを行うというプロセスとなっている。

このプロセスで重要な点の一つは適切な学習データを AI に学ばせる必要があるということである。SDQ モデルの構築には過去のデータとクエリが用いられ、クエリ解決前後のデータパターンから学習データを用意している。図 1-2-2 はパターン認識のイメージを表している。

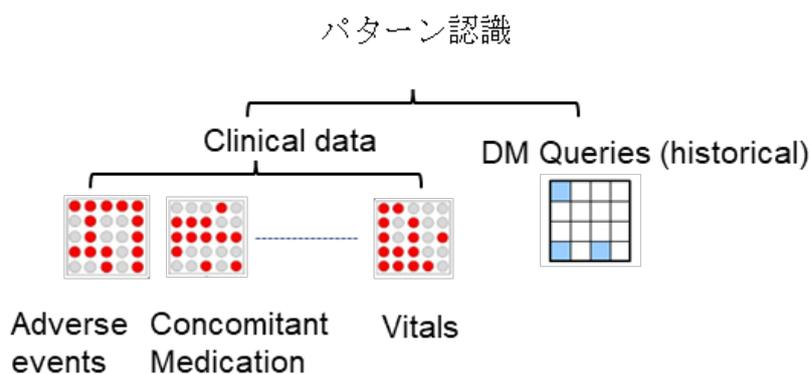


図 1-2-2

SDQ プロセスには以下の要素が含まれている。

- 過去の臨床データとそれに対するデータマネージャー発行クエリから AI が学ぶ機械学習モデルの開発

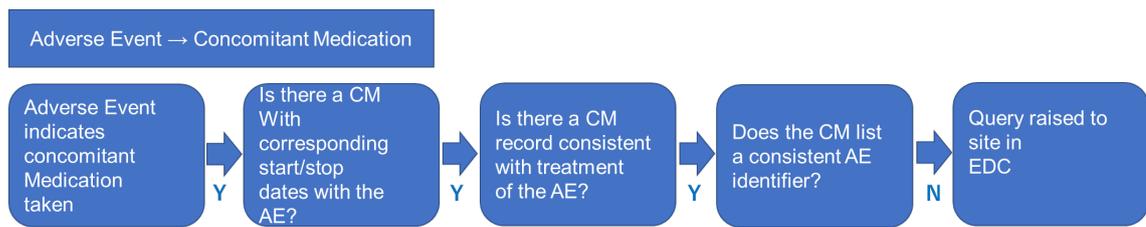


図 1-2-4

また、「併用薬の投与理由が有害事象とされている」場合、「その併用薬投与に対応する発現日/消失日を持つ有害事象があるか？」⇒「その併用薬による治療と整合する有害事象はあるか？」⇒「その有害事象は薬剤による治療を受けているか？」という順に質問をしていき最後に初めて「いいえ」であった場合に、SDQはこのシナリオの「齟齬の分類」に該当する齟齬ありと判定し対応するクエリを生成する（図 1-2-5）。

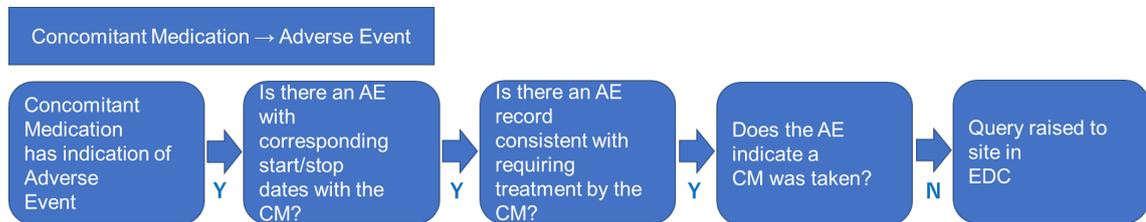


図 1-2-5

上の「Adverse Event → Concomitant Medications」のチェックの例の場合に、自動生成されるクエリの事例を下に示す（図 1-2-6）。

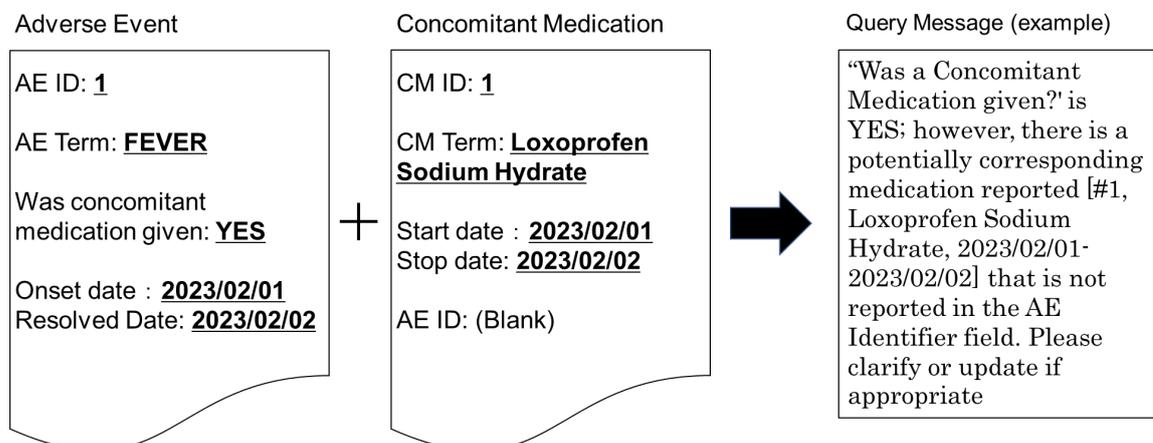


図 1-2-6

次に、SDQに使われている Teach cycle という方法論について説明をする。このサイクルはまず「齟齬の分類」を定義するところから始まる。SDQでは齟齬を検出するための複数のモデルがあり、どのモデルでどのような齟齬を検出するかを定義している。そしてモデルに学習データとなる臨床データとクエリが与えられモデルが構築される。このようにして構築されたモデルは、人間による予測に対する採否のフィードバックを受け、それによっ

て再度学習することになる。更に偽陽性、偽陰性の評価結果から必要に応じて、学習データのバージョンアップやモデルのバージョンアップを行うことにより、その予測精度は向上していくことになる（図 1-2-7）。

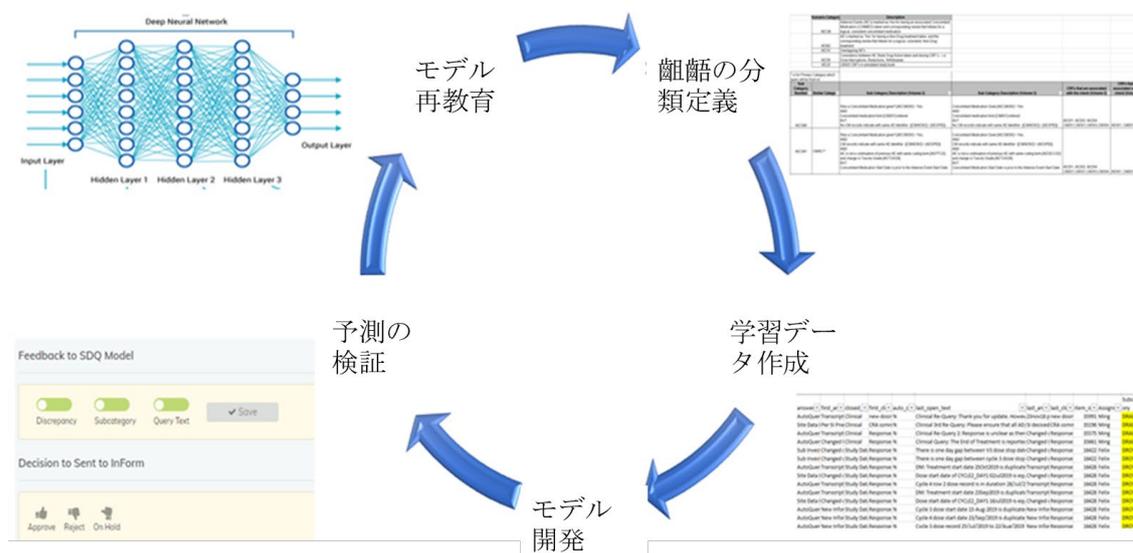


図 1-2-7

SDQ 導入の価値は、データベース固定などのマイルストーンの早期達成のためにデータレビュープロセスを加速させることにある。Pfizer 社で実施した新型コロナワクチンのある国際共同治験では、従来約 27 日かかったクエリ解決の期間が 1.7 日に短縮でき、それが短期間にデータベース固定を達成する上で大きなインパクトを与えたとコメントしている。クエリ解決までの期間の短縮に SDQ 以外の要因による貢献を考慮しても、SDQ による効果は非常に大きいものと考えられた。

SDQ によりもたらされるベネフィットを以下に示す。

- ・ 最後の参加者の最後の来院からデータベース固定までの時間短縮
- ・ データ整合性確認のためのレビューおよびクエリ文作成に要する労働力の削減
- ・ データ間の整合性を含む試験データの品質の改善
- ・ ルーチンクエリおよび複雑なクエリの自動化
- ・ リアルタイムに近いクエリ発行
- ・ (human-in-the-loop を伴う) AI によるクエリテキストの自動生成

1.3. データレビューならびにクエリ文作成②

EDC データに対して、データチェックを実施した結果、クエリ発行が必要と判断された場合に、クエリを発行する。この一連の工程において、チェック結果確認からクエリ発行作業は、人が判断し実施されている。

本節では、クエリ発行を、教師あり機械学習で実施した Pfizer 社以外の事例を紹介する。複数の条件からクエリ発行を行う場合、その工程が複雑であれば、目視でのチェックを行ってきた。この複雑なチェックもランダムフォレスト（図 1-3-1、Appendix1.1.1 参照）のような機械学習の手法で、コンピューターがクエリを発行する判断を行うことが出来る [3]。

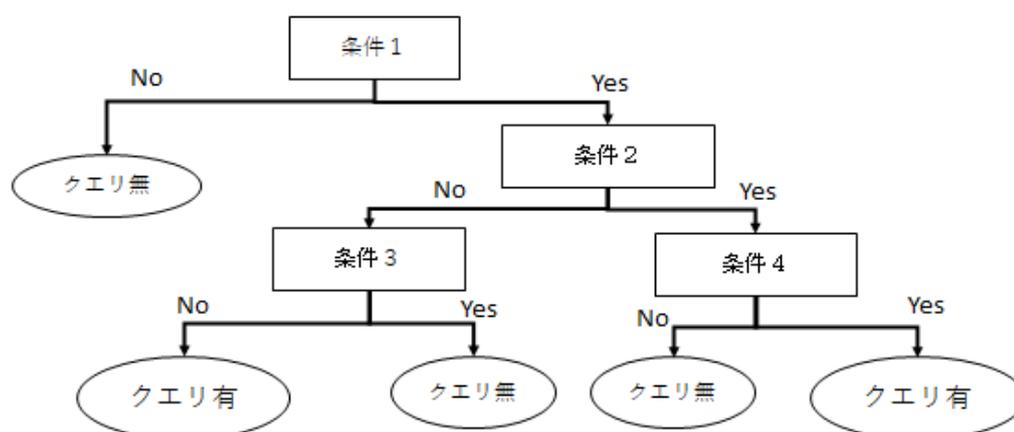


図 1-3-1

初めに、学習用データとして、過去の試験データ等にクエリ発行有、クエリ発行無のフラグを付けたものを作成する。このクエリ発行有、クエリ発行無のフラグは、データマネージャーが判断してつけたものである。

なお、学習データと実際のデータでデータベースの構造が異なると判断できなくなるため、全てのデータを SDTM 形式等にそろえておくとよい。

次に学習データをランダムフォレストを用いて学習させ、実際のデータに「クエリ発行有 or 無」のフラグを付ける。最後に「クエリ発行有」のフラグがついたデータにクエリ文章を表示させ、クエリを発行する（図 1-3-2）。

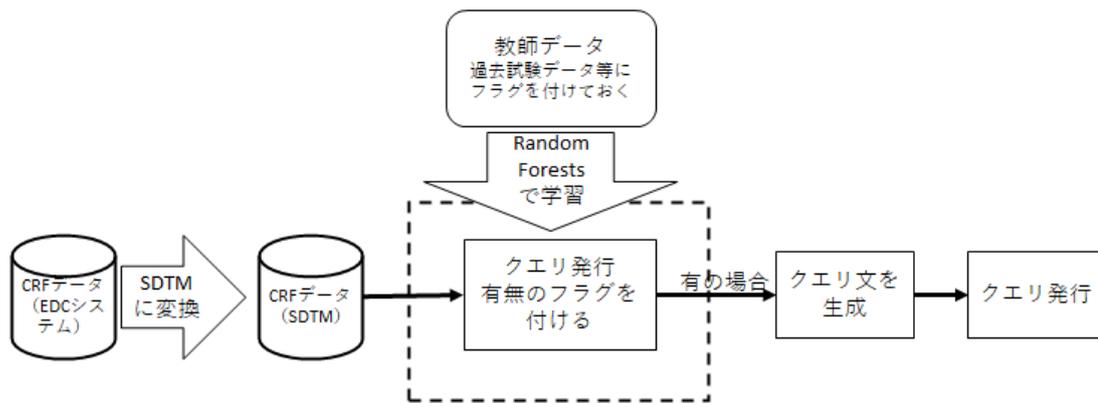


図 1-3-2

目視チェックでは、併用薬のチェックであるような添付文書の内容とのチェックも行われているが、添付文書のデータを学習させることで、そのようなチェックも実行可能と考える。

2. AIリテラシー向上・学習データの準備・バリデーションについて

AIの導入に際して、さまざまな問題や懸念が生じる可能性がある。ここでは、AI導入の際に共通する問題点や課題として、AIリテラシー向上の重要性、学習データの準備に関する課題、およびバリデーションについて取り上げる。

2.1. AIリテラシー向上の重要性

デジタル人材の育成に関しては、内閣府、文部科学省および経済産業省において、「数理・データサイエンス・AI教育プログラム認定制度（リテラシーレベル）」[14]が創設されることとなり、文理を問わず全ての大学・高専生が、正規課程にてリテラシーレベルの数理・データサイエンス・AIを習得することを目標とし、大学等の卒業単位として認められるようになる。

『「数理・データサイエンス・AI教育プログラム認定制度（リテラシーレベル）」の創設について』[15]の資料に示されているように国としてITリテラシーを向上させる取り組みがなされている（図2-1-1）。デジタル化が進む現在においては、AI業務に携わらない人材に関してもリテラシーレベルの知識が、必要と考えられる。AI業務に携わる人材に関しては、AIの活用策を検討したり、外部ベンダーにAIを使用したシステムの導入・維持管理を委託する場合においても、ベンダーへの委託内容の検討や、ベンダーに対する評価などで、AIに関する応用基礎レベルの能力は必要になると考えられる。データマネジメントに携わる者も、個人レベルでより一層自己啓発していく必要があり、また組織としても個人の必要

なスキルセットとして学習を促していく必要がある。

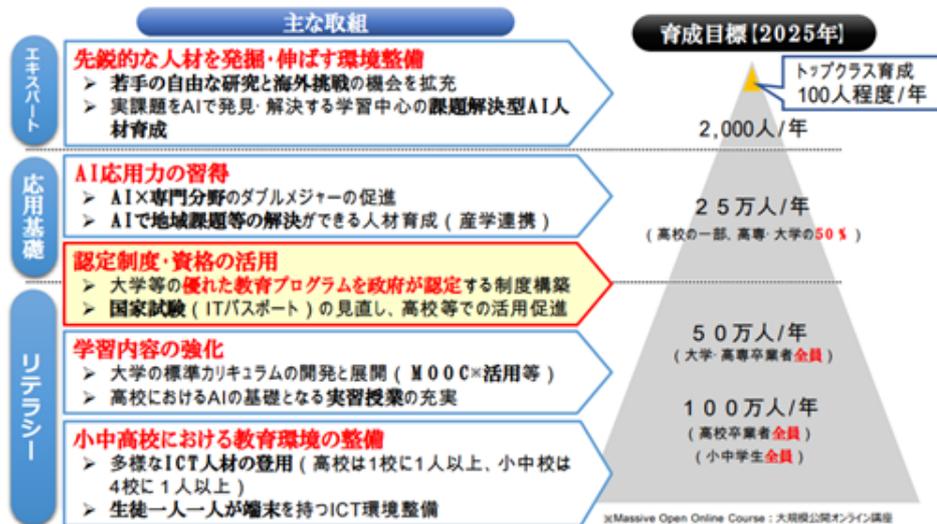


図 2-1-1

2.2. 学習データの準備に関する課題

AI 学習に必要なデータに関して、十分な量の学習データが必要になるが、データ量に依存して精度が高くなるわけではない。少ない学習データで高い精度が得られたとしても、未知の新しいデータでは精度が得られないといった問題や、学習データが大量にあっても似通ったデータが多かったり、データに多様性がない場合も、新しいデータでは精度が得られないといった問題がある。学習データの数については、2020年の製薬協の成果物「機械学習始めました」の第2章第1節に、AIプロジェクトで使用されたデータセットの推定規模に関する記載があるため、必要に応じて参照いただきたい。

一般的に、機械学習で十分な認識精度を実現するには、事前に大量の学習データを用意し、AIに学習させなければならない。ここではデータ齟齬の検出のためにAIを利用する場合を例に上げて、学習データ準備の課題について述べる。過去試験のデータを学習データとして使用しようとする、アーカイブされているのはレビュー済みのデータであるため齟齬のあるデータの件数は少なく、AIで検出したい十分な齟齬のあるデータのサンプル数が含まれていない事態が起こり得る。その場合は、別途、齟齬のあるデータの用意のためのリソースや時間を確保するなどの対応が必要となる。

DM分野に限らず、AIにおける課題として、学習データの準備や、大量データによるコンピューターへの負荷などが問題視されているため、これらの問題に対する新たな技術や手法は、今後も増えていく事が予想される。

データの収集や加工にも大変な時間と労力を要するため、まずは PoC を実施して必要な

データ量と質を見積ることや、ある程度学習モデルの開発が進んでいるモデルに対しては新たなデータ追加が学習モデルの精度を向上することに寄与するかを評価することが推奨される。

データ量の学習結果（精度）への影響は用いる指標、および手法（アルゴリズム）によっても異なる事が知られている。変数の少ない構造化データに対して、シンプルな手法を用いた場合のシミュレーション [16]では、比較的少ないデータで性能が頭打ちとなるが、より複雑なデータや、画像など非構造化データを用いる深層学習に分類される手法の場合に性能を向上させるためには、指数関数的にデータを増やす必要がある点について言及されている。

2.3. システムバリデーション

これまでデータマネージャーがデータレビューの際に実施したプログラムチェックでは、同じ **Input** に対してプログラムチェックを実施した場合には常に同じ **Output** が出力されていたため、バリデーションも容易に実施することができた。AI を使用したシステムでもシステム導入時であれば、“Locked” approach（内部の設定を固定することで、同じ **Input** を与えると同じ **Output** が出力される）が使用できるかもしれない。しかし、AI を使用したプログラムやシステムでは毎回同じ **Output** が出力されるとは限らない。例えば、システムのプログラムが読み込んだデータにより常にアダプティブに変更される場合、毎回 **Output** は異なってしまう。また、出力の精度はデータの量や品質に依存してしまう。

このようなAIを使用したシステムに対しては、バリデーションにリスクに基づくアプローチの考え方を導入することも有効な手段と考えられる [17]。あらかじめ、バリデーション計画書にバリデーション方法について定義しておき、適宜チェックすることでバリデーションの記録として用いることができるかもしれない。

バリデーション計画書に記載すべき項目

- バリデーションの目的、スコープ、各担当者の責任範囲とアクション
- バリデーションの方法（例えば、**Output** の適切性を確認）
- 学習データの適切性
- バリデーション実施時の Go/No-go の閾値
- フィードバックプラン
- データの品質によるリスク及び解決方法

3. 規制・倫理・法律上の課題

AI の導入が各分野で進む中、その基盤となる AI 原則については、医療分野に留まらない世界全体の共通認識が確立されつつある [18]。一方、AI 原則に基づく規制の整備の進捗度合いは分野ごとに大きく異なっており、医薬品に限定すると、世界を見渡しても明確な規制

要件が存在しないのが実情である。そのため、本書のテーマであるデータマネジメントでの AI の利活用を検討する場合、まずは分野横断的に示されている AI 利活用に係る原則や要件等を把握し、最低限それらに留意することが求められる。

本章ではまず AI 原則を始めとする AI に係る分野横断的な規制確立に向けた動きについて紹介した後、AI に係る医療分野での規制の現状を概観する。その上で、データマネジメントにおける AI の利活用にあたり特に留意すべき事項について、倫理、法律および説明責任の観点から記述する。

3.1. AI原則、AIに係る規制確立に向けた動き

これまでに世界各国の政府に加えて各種ステークホルダーや企業から AI 原則が公表されており、日本でも、AI が社会に受け入れられ適切に利用されるための原則として、2019 年 3 月に「人間中心の AI 社会原則」が内閣府から示された [6]。更に 2019 年 5 月には、日本を含む全 42 か国間で合意された AI 原則が OECD（経済協力開発機構）から公表されるなど [19]、国際的なコンセンサスが形成されつつある。

「人間中心の AI 社会原則」には、社会（特に、国などの立法・行政機関）が留意すべき AI 社会原則として、①人間中心の原則、②教育・リテラシーの原則、③プライバシー確保の原則、④セキュリティ確保の原則、⑤公正競争確保の原則、⑥公平性、説明責任及び透明性の原則、⑦イノベーションの原則、の計 7 つの原則が示されており、これは OECD から公表された AI 原則と一致した内容となっている [20]。更に、AI の研究開発と社内実装に従事する開発・事業者側は、「AI 開発利用原則」として上述の 7 つの原則を踏まえた AI 利活用のための原則を定め、遵守すべきとしている。

更に欧州では、欧州委員会（EC）が 2020 年 2 月に「AI 白書 [21]」を発表し、市民の価値観と権利を尊重した安全な AI 開発の「信頼性」と「優越性」を実現するための政策オプションを示した。また翌年 4 月にはこの内容を具現化し、AI 規制枠組み規則案等を含めた「AI 政策パッケージ [22]」を発表するなど、より具体的な規制確立に向けた動きが見られる。

3.2. AIに係る医療分野での規制の現状

AI 利活用に当たっての分野横断的なゴールが AI 原則として確立されてきた一方、その原則と各社の AI ポリシーや組織ガバナンスの間に位置するガイドラインのような中間的なルールは、特定の用途に対する限定的なものを除き、未だ策定が求められている状況である [18]。具体的に従うべき規制が明確でないために、いざ AI の利活用を検討し始めても多くの分野で手探りにて進めざるを得ず、これは医薬品分野においても例外ではない。

医薬品に限定せず医療分野全体を見渡すと、AI を搭載した医療機器に対するガイドラインの整備が海外で先行して進んでいる。具体的には、2021 年 10 月に米国食品医薬品局（FDA）がカナダ保健省、英国医薬品医療製品規制庁（MHRA）と共同で「Good Machine Learning

「Practice for Medical Device Development: Guiding Principles」を発行し [23]、AI と機械学習を使用する医療機器開発にあたっての指針を示している。PMDA からは上述のような指針が示されていないものの、AI を搭載した医療機器開発に際しての相談窓口（医療機器プログラム総合相談）を 2021 年 4 月に設置し、審査体制の強化が行われた。また、画像診断支援システムなど、特定の用途に対する限定的な規制については国内外で随時整備が進んでおり、並行して AI を搭載した医療機器の開発事例が増加している [24]。

一方、執筆時点において、医薬品に特化した AI 利活用の要件を明示している規制当局は、世界全体を見渡しても存在しない。これは、現在医薬品の研究・開発過程で利活用されている AI の事例は、創薬における候補化合物の調査や安全性監視におけるシグナル検出など、探索的側面を持つ領域での使用が大多数であり、意思決定に直接関わる場面での利活用事例が少ないことが一因と考えられる。また、2021 年 7 月に経済産業省から出された報告書「我が国の AI ガバナンスの在り方 ver. 1.1」の中で、医療分野においては、これまでの規制に対する考え方や設計思想を活かし、業法を所管する組織によるルールメイキングを尊重することが望ましいと述べられている [18]。従って、日本で医薬品に特化した AI 利活用に関する規制を確立するためには、各企業やステークホルダーがその経験を元に規制当局と協同でその整備を推進していく必要がある。

3.3. 倫理上の留意事項

AI による判断は、設計や学習に使用するデータの内容に大きく左右される。そのため、偏ったデータを学習に用いた場合、AI から導き出される結果にはバイアスがかかり、人種や性別、国籍、宗教等の特定の集団に対する偏見の助長に繋がったり、プライバシーへの配慮の欠けた結果を招く恐れがある。Amazon 社の採用プロセスに用いた AI ツールが、過去の応募者の大半が男性であったことを背景に、男性を積極的に採用したことなどが例に挙げられる。

AI に人間と同様の倫理観を持たせる、あるいは AI を使用してもなお適切な倫理観を保った環境とするための対策として、大きく 3 つ挙げられる。

1 点目は、学習に用いるデータによるバイアスへの留意である。まずは学習データ自体にさまざまなバイアスが存在することを認識し、それが社会的に許容されるか否かを判断した上で、可能な限り偏りのない代表性¹のあるデータを用いることが求められる。また教師あり学習の場合、学習データに対して人間が付与するラベルにもバイアスが生じうることから、例えばラベリングの統一的な基準の作成などが有効と考えられる [25]。

2 点目は、学習アルゴリズムによるバイアスへの留意である。特に、機械学習では多数派が尊重され少数派が反映されにくい傾向がある（バンドワゴン効果）。そのため、バイアスの要因となりうる個人の属性を明確にした上で、学習時のモデルの予測因子から該当する属性を除外する、該当する属性に左右されない結果が導出されるようモデルを調整する、な

¹ サンプルとして抽出され利活用に供されているデータが、その母集団の性質を正確に反映している度合いのこと

どの対応が求められる [25]。また、学習データとテストデータの間で独立性を確保することや、運用開始後の定期的なモデルパフォーマンスの見直し時に、構築時には予期していなかったバイアスが生じていないか、確認すべきとしているガイドラインも存在する [23]。

最後に 3 点目として、人間の判断の介在が挙げられる。前述の 2 つの対策を講じてもなお、バイアスの完全な除去は困難なことから、AI による判断を基盤に、最終的な決定や判断は適宜人間が行う手順を組み込む (Human-in-the-loop) ことも必要とされる。またこの対応は、倫理面の配慮を可能とするだけでなく、AI を利活用する業務における判断の責任の所存を明瞭化するためにも有効な手段である。なお「AI 白書」では、リスクに基づくアプローチの考え方を取り入れ AI をリスクに応じて分類した上で、医療分野などで使用されるハイリスク AI²に該当する場合、人間が事前に確認や妥当性評価を行っていない AI を使用したシステムが導き出した結果は十分な信頼性を持たないとしており、運用時においても人間による適切な関与の確保が必要と述べられている [21]。

3.4. 法律上の留意事項

3.4.1. 概論

大概の場合、AI の開発、運用の過程では個人情報を含むデータを扱うこととなるが、AI 原則に含まれる通り、その際はプライバシー確保への対応が必須となる。また、AI を使用したシステムの開発、運用にあたっては、社内にリソースを十分に持たない多くの企業が外部の企業と共同で進めていくこととなる。その過程でさまざまなデータやプログラムなど多種多様なリソースの使用や成果物の作成が発生するが、それらに係る権利などの法律的側面についても最低限の理解と留意が必要となる。

知的財産権 (著作権、特許権等) や個人情報保護法による保護内容は国ごとに異なる。また、例えば著作権に関して、具体的にいずれの国の法が適用されるかは裁判所等の判断に委ねられるが、著作物の利用行為や著作権の侵害行為については、これら行為の行われた地の法により規律されることが少なくないとされている [26]。そのため、海外のサーバーへの生データ等の保存や海外でのデータ加工を行う場合等においては、サーバー設置地や加工を行った地の著作権法への留意が合わせて必要となる。以降の項では AI の開発および運用が全て日本国内で完結する場合に焦点を当てて、本邦の法律に関し記述するが、上述の通り海外にも跨った利用がなされる場合は、該当国の法律に留意する必要があり、本書に記載の限りでないことに留意されたい。

データマネジメント業務での AI 利活用を検討する際、特に留意が必要な法律としては著作権法、特許法、個人情報保護法の 3 つが挙げられる。どの法律に留意すべきか、ならびにどのリソースや成果物がその考慮対象となるかについては、AI を使用したシステムの開発

² セクターおよび利用方法の 2 つのクライテリアに基づき、重大な物理的・精神的ダメージを及ぼす蓋然性が高いもの
・セクター：医療、交通、エネルギー
・利用方法 (事例)：個人や企業の権利に法的効果または同様に重大な影響を生み出す、傷害・死亡・重大な物質的・非物質的損害のリスクを呈示する、個人や法人が合理的に回避することのできない影響を生み出す

段階³により変化するため、他社と秘密保持契約書等の各種契約を結ぶ際は、どの開発段階を前提とした契約であるか両社間で明確にした上で、開発過程での使用物や成果物を整理、必要な法律を考慮し、契約に含めることが有用となる。例えば、AIの導入可否の検証段階においては、モデル構築の際サンプルデータのみが使用されることが多く、その場合は個人情報保護法に対する考慮が不要となる。

なお、「3.4. 法律上の留意事項」は、各種文献からの引用を主体として記載している。各社でAIを実装される際は、本書の内容を踏まえつつも改めて各社の法務部門への相談が必要であることにご留意いただきたい。

3.4.2. 著作権法、特許法

著作権法については、創作性を持つ可能性のあるデータ（生データ、学習データ、学習済みパラメーター、入力データ、AI成果物等）およびプログラム（学習用プログラムや推論プログラム等）のソースコード部分はその考慮対象となりうる。

特許法については、新規性及び進歩性等の要件を満たす可能性のあるプログラムのアルゴリズム部分及びノウハウ⁴がその考慮対象となりうる。

これら知的財産権に係る法律として包括される著作権法及び特許法については、契約書にて成果物ごとの権利の帰属先を明確にすること、ならびに成果物の利用条件を明記し、それに従うことが対応として求められる。また、知的財産権の対象とならない成果物についても、その利用等に関する法律上の明確なルールが存在しないため、契約により当事者間の権利関係を定めあらかじめ合意しておくことが推奨される [26]。

例えば生データや学習データは、著作権法上の「データベースの著作物」⁵に該当する場合にその保護対象となるが、該当する事例は多くはないとされている [27]。そのため一般には、データに適法にアクセスし、その利用をコントロールできる事実上の地位、または契約によってデータの利用権限を取り決めた場合にはそれに基づいてデータの帰属先を決めることとなる。また、プログラムに含まれる学習済みパラメーターについては、大量の数値データであるためその創作性等が認められず、通常は著作権など知的財産権の対象にはならない可能性が高いとされている [26]。

³ アセスメント段階、PoC 段階、開発段階、追加学習段階

⁴ ベンダーにおいては、AI技術の研究・開発に関して有する知見、技術、情報等が、ユーザにおいては、AI技術の適用対象となる業務に関して有する知見、技術、情報等が含まれる。以下に係るノウハウが例として挙げられる。

- ・生データの取得・選択
- ・学習データへの加工（ラベル情報の付与等）
- ・学習用プログラムを用いた学習（ハイパーパラメーターの調整を含む）
- ・学習済みモデルの調整

⁵ 著作権法（第一章 第一節 第二条 十の三）では、「データベース」の条件を以下としている。

- (i) 論文、数値、図形その他の集合体であること
- (ii) (i)に係る情報をコンピュータを使って検索できるように体系的に構成していること

また、「データベースの著作物」の条件は以下としている。（同法：第二章 第一節 第十二条の二）

- (i) 集めたデータがコンピュータにより検索可能な状態になっていること
- (ii) データの情報選択または体系的な構成について創作性があること

また、データマネジメント業務に係る AI を使用したシステムの場合、使用する生データや学習データには MedDRA や WHODD 等ライセンスが関係するデータが含まれる可能性も高い。社内利用に留まる AI を使用したシステムの場合は特段問題にならない事項と考えられるが、一方で他社と該当データを含むやり取りが発生する場合、更なる注意が必要である。具体的にはまず、上述のような辞書を用いたコーディングデータの場合、辞書から必要なデータを選択し付与した、という観点から、データの情報選択について創作性を持つ「データベースの著作物」として著作権による保護対象となる可能性があり、それに応じた対応が必要になる。また、データの受け渡しが発生する各社が該当のライセンスを保持しているか、ならびに該当のライセンスを使用し作成されたデータの利用方法が、ライセンス利用規約上の利用目的の範疇に収まっているかを確認し、場合によっては、データから一部情報を削除する対応も必要と考えられる。

3.4.3. 個人情報保護法

個人情報保護法については、主にデータ（生データ、学習データ、入力データ等）が考慮対象となる。データマネジメント業務にて使用される生データは、参加者から収集した臨床試験データが大半を占めると考えられるが、当該データは、被験者識別コード⁶を使用している場合でも個人情報でないと切り切れないことに留意する必要がある [28]。個人情報保護法の対象となるデータを使用する場合、主な対応としては、契約書へ個人情報保護法に則った適切なデータ加工又は同意取得等の手続きを履行すること、および各種データの利用条件を明記した上で、契約書の記載に従った対応の実施が求められる。なお、過去に取得済みのデータであり同意取得が困難な場合であっても、公衆衛生例外規定に該当⁷し、更に一定の条件⁸を満たす場合を除き、同様の対応が求められることに合わせて留意されたい。

同意取得について、個人情報に該当する臨床試験データの場合、基本的には要配慮個人情報に該当することから個人情報保護法上のオプトアウト規定⁹を適用出来ないため、加工せず用いる場合は、データの二次利用として同意書への記載が必要となる。また、データ加工により同意取得が不要となる場合も、加工によりプライバシーが守られることなどについて可能な範囲で説明をすることが望ましい。一方 PMS データの場合、GCP で管轄される

⁶ 個々の被験者の身元に関する秘密を保護するため、治験責任医師が各被験者に割り付けた固有の識別番号

⁷ 公衆衛生の向上又は児童の健全な育成の推進のために特に必要がある場合であって、本人の同意を得ることが困難であるとき

⁸ 主に次の 3 点の条件が適用できる場合に限り利活用できるとされている

①「製薬企業が行う有効な治療方法や薬剤が十分でない疾病等に関する疾病メカニズムの解明、創薬標的探索、バイオマーカー同定、新たな診断・治療方法の探求等の研究」といった公衆衛生の向上に特に資する場合

②自社内の研究のために用いる場合

③学会、論文、ホームページ等において結果を公表することが前提にある場合

⁹ 個人情報保護法上のオプトアウトと医学研究におけるオプトアウトは、全く異なる制度を指すことに留意されたい。個人情報保護法においてオプトアウトという用語は、第 27 条第 2 項に基づき、本人の求めに応じて当該本人が識別される個人データの第三者への提供を停止することとしている場合、定められた事項について、あらかじめ本人に通知し、または本人が容易に知り得る状態に置くとともに、個人情報保護委員会に届け出たときは、あらかじめ本人の同意を得ることなく、当該個人データを第三者に提供することができる制度（オプトアウトによる第三者提供）を意味する。

製造販売後臨床試験とそれ以外（使用成績調査、特定使用成績調査）で分けて考える必要がある。前者の場合、前述の臨床試験データと同様の考え方が必要と考えられる。後者においては、そもそも実施に際して同意取得が必須となっていないこと等から、データの二次利用に関する同意取得や説明の必要性について規制上明確になっていない。

データに対し加工作業を行い匿名加工情報または仮名加工情報とした場合、前者は本人同意なく第三者提供が可能な一方、後者は、原則第三者（業務委託先、事業継承、共同利用の場合を除く）へのデータ提供は不可¹⁰となる。反面、仮名加工情報の方が求められる加工の程度が小さく、故にデータの情報量が低下しにくい利点を持つ。データ加工を行う場合は、いずれの情報とするか、各情報の特性を踏まえて検討した上でガイドラインに従った加工作業が求められる。

各情報の特性やデータ利用にあたっての留意事項等の詳細は、2022年4月に日本製薬工業協会 医薬品評価委員会 臨床評価部会から公表された「医薬品開発及びデータ二次利用における個人情報保護に関する留意点 [28]」も参照されたい。

3.5. AI利活用に伴う説明責任について

AI利活用に際しての明確な規制やガイドラインが定まっていない中、AIを使用したシステムを開発、利用する各企業にどこまでの説明責任が求められるのか把握が難しい状況である。一方、人間中心のAI社会原則では、「公平性、説明責任及び透明性の原則」の中で、AIより導き出された結果に対する説明責任が適切に確保されること、ならびに技術（AIとそれを支えるデータやアルゴリズム）に対する信頼性が担保されることを必要としている[6]。

具体的にどの程度の対応が必要となるか悩ましいところであるが、総務省から公表された「国際的な議論のためのAI開発ガイドライン案 [29]」や「AI利活用ガイドライン [25]」において、説明可能性確保のための手段として、以下が例示されている。なお、上記原則はアルゴリズムやソースコード、学習データの開示を想定するものではないとしている。

- 1) 可読性の高い、解釈可能なアルゴリズムを実装するAIソフトの採用
- 2) 解釈可能性の低いアルゴリズムによる判断結果の説明を行うための技術的手法の採用
- 3) AIの学習等に利用されたデータの来歴（いつ、どこで、どういう目的で収集されたものか）の管理
- 4) AIに対する複数の入力と出力の組み合わせに基づく、学習モデルの判断の傾向の分析（例：入力パターンの変化による出力の変化の観測）

このうち1)2)については特に内容が抽象的であり、実務として必要な対応を把握しにくい部分があるが、これについては2020年5月に英国のICO (Information Commissioner's Office) から発行された「Explaining decisions made with AI [30]」が参考になる。当該文書

¹⁰ ただし、作成元の個人データは本人同意があれば提供可能

は、AIに基づく決定（AIからの出力結果及びそれに基づく人間の判断介入後の決定）に関する説明を行うにあたっての Good Practice の実務指針を示したものであり、その中で解釈可能性の高いアルゴリズムと低いアルゴリズムが例示されている。また、後者を使用したモデルをブラックボックスモデルと定義し、使用すること自体は許容出来るとしている一方、条件として利用前に潜在的な影響やリスクの徹底的な検討、LIME（Local Interpretable Model-Agnostic Explainer）を始めとする補助的な解釈ツールの使用などが必須になるとしている。更に当該文書では、説明責任を十分に果たすための組織内での各部署の役割分担、ポリシー等の規則、記録や文書管理の在り方についても言及されており、AI 利活用と並行して、これら観点からの組織体制の整備も必要であることが認識出来る。

4. AI導入検討方針および投資

4.1. AI導入検討の方針

経済産業省はAIに対して「AIが何かがわからない、メリットが想像できない」、「自社に導入できるのかわからない、自信がない」、「まず何を準備して、どうやって進めればいいのかわからない」といった不安を持っている中小企業のために複数の「AI導入ガイドブック」を作成している。このガイドブック「構想ステージ」編では、AIとは何か、導入のメリット、AI取組領域の選び方、AI導入に向けて準備すること等が簡潔に述べられていて [31]、これからAI導入の検討に取り組もうと考えているデータマネジメント組織にも有用と思われる。

この「構想ステージ」編ではAI取組領域検討の3つのステップとして以下を挙げている。

ステップ1：自社の経営目標・業務課題の把握

0. 経営目標の明確化
1. 業務プロセスの見える化
2. 現場へのヒアリング、課題抽出
3. データの有無と内容の確認

ステップ2：AIを用いた解決案を幅出し

1. 課題に対応するAIソリューションの調査
2. 1. が自社で技術的に実現可能かを調査
3. 皆でアイデアを持ち寄り、フリーディスカッション

ステップ3：解決案を評価して優先順位付けし、取組領域を決定

1. すぐできそうか（実現難易度）を評価
2. できたらどれだけメリットがあるか（得られる効果）を評価
3. 評価した解決案を優先順位付けし、取組領域を決定

経営層に対しては、AI導入および組織変革のためには、経営者自らが推進することに加え、推進者を任命すること、推進者が検討・推進に取り組みやすくなるような配慮が必要で

あると述べている。そして従業員に対しては、『成果に繋がるか』よりも、AI で何ができるだろう?まずはやってみよう!という姿勢で検討を進めていくことが重要』と述べている。また、「AI 導入の心構え」としては、「実務への AI 導入にあたっては、将来達成したいビジョンに向けて『小さく・素早く始め』、『継続的・段階的に育てていく』心構えが重要』と述べている。将来達成したいビジョン・価値を見据えて限られた業務範囲での AI モデル構築・検証から始める方針である。

Forbes Business Council メンバーは、組織内で AI を採用および実装するには、相当の時間、労力、費用、および専門家のサポートが必要となることから、中小企業の経営者がビッグデータやプログラマーなしで AI を採用する際のアドバイスとして以下のような注意点または留意点を挙げている [32]。

- ・ コストベネフィット分析を実施する
- ・ 既存の技術を最大化する方法を学ぶ
- ・ 実験を実施しておき、将来どこに投資するかを決定する
- ・ サードパーティのプラットフォームを活用する
- ・ 誇大広告を見抜く (AI と言っているが実際には RPA 等である場合もある)
- ・ 専門家に相談する
- ・ コーディングの必要のない手頃な価格の AI 製品を探索する
- ・ 改善または強化を目指している領域の全体像を把握する
- ・ プラグインとオープンソース AI データハブを探索する
- ・ 技術パートナーと協力する
- ・ 他のソフトウェアと合わせて提供されているソフトウェアソリューションを利用する
- ・ 完全に準備ができたときにのみ採用する

実際には、Forbes Business Council メンバーの挙げている点に留意しつつ、続くセクションで述べる投資対効果の評価を行いながら、経済産業省が中小企業に示しているような方針で AI 導入検討を進めることになると思われる。

4.2. AIに対する投資対効果の評価

さまざまな分野で AI 導入による省力化、効率化、品質改善等が期待されており、臨床試験のオペレーションにおいてももちろん例外ではない。製薬企業においても AI 活用を今後の重要戦略に掲げている部門は少なくないだろう。しかし AI は何かを達成するための手段であって AI 導入そのものが目的ではないはずである。投資が必要な AI の導入を進めていくためには、投資に見合うベネフィットが得られるまたは得られていることをマネジメント層に示すことは必須である。

しかし、AI によるベネフィットを数量的に示すのは容易ではない。人が行っているプロセスを完全に AI が代わりに行うことにする場合などは、人件費や委託費用から比較的容易

にベネフィットを金額として算出できるであろう。一方、品質向上のような数値化しにくいものがアウトカムである場合、またある問題が発生する確率を AI によって低減させる場合などは、ベネフィットの算出が難しくなる。特にソリューションとしての AI の導入の初期段階では、AI の予測の偽陽性や偽陰性の可能性が高い場合も想定され、そのために人が AI の予測を確認することを考慮しなければならないことも考えられるため、AI の導入によりむしろリソースがかかるという場合もあり得る。また、AI 導入前後で何らかのメトリクスに差があったとして、それが AI によるものなのか、それ以外の要因に起こったものなのか明確にできない場合もあり得るだろう。

AI によるベネフィットの数量化については、以下のステップで検討する方法が紹介されているので [33]、身近な例を交えながらこのステップに沿って説明する。

- ステップ 1 : ビジネス上の成功指標を定義する
- ステップ 2 : データによってアウトカムを測定する
- ステップ 3 : アウトカムと AI の関係を確立させる
- ステップ 4 : 全てのコストを考慮して投資対効果を計算する

ステップ 1 : ビジネス上の成功指標を定義する

AI 導入にあたっては、Lead Indicator（先行指標）と Lag Indicator（遅行指標）を特定することにより、AI 導入のアウトカムを明確にする必要がある。Lead Indicator は Lag Indicator に先行して発現する指標であって、最終的なアウトカムを示す Lag Indicator を左右する。

例えば、最後の患者の最終来院からデータベース固定までの期間を短縮するというデータマネジメント組織に課せられた課題を解決するために AI を活用することにしたとする。その場合、個々のクエリ発行に要する期間が Lead Indicator となり、最後の患者の最終来院からデータベース固定までの期間が Lag Indicator となる。Lead Indicator の改善が Lag Indicator の改善につながり、Lead Indicator の改善なくして Lag Indicator の改善はないからである。

ステップ 2 : データによってアウトカムを測定する

アウトカムを定義したら指標を測定するメトリクスを定め、必要なデータを収集する。この際に既に収集しているデータに限定することなく必要なデータを収集することが極めて重要である。先の例であれば、これまで個々のクエリ発行に要する期間、すなわち日数のデータを収集していなければ収集し、また収集したとしても評価していなければ評価の対象とすることになる。

ステップ 3 : アウトカムと AI の関係を確立させる

次のステップは収集したメトリクスによってアウトカムに対する AI の寄与を明確にする

ことである。これは薬剤投与における有効性の評価にも似ているが、一方、これには有効性評価において薬剤投与以外の影響を取り除くことの難しさと同様の難しさが含まれている。薬剤投与における有効性の評価であれば、プラセボ対照二重盲検試験という手法を用いて薬剤効果以外の要因による効果を相殺して薬剤の効果を明確にすることもできるが、AIの寄与を測るために同様の手法を用いることはあまり現実的ではないだろう。その場合、過去のメトリクスをヒストリカルコントロールとして使用してAIの寄与を考察しなければならない場合も十分に想定される。ステップ1で取り上げた例であれば、過去のLead IndicatorとLag Indicatorについて単位期間あたりの収集データ量、発行クエリ数、投入リソース量等を考慮して調整した上でAIの効果を見極める必要がある。

ステップ4：全ての費用を考慮して投資対効果を計算する

AI導入のために必要な初期設備投資、人件費も含めた開発費用、ユーザートレーニングに要する費用およびリソース、システムメンテナンス費用に加えて、場合によっては外部データのライセンス費用なども含め全ての費用を計上することが重要である。その上でAIによって実現しようとしているアウトカムがAI導入や運用のための投資に見合うかの判断をする必要がある。アウトカムを金銭的に評価することは必ずしも容易ではないし、患者に対する安全性や組織の社会的信頼など、金銭的に評価できない場合もある。

5. AIに取り組むためのヒント

5.1. 概説

AIで解決できそうと思える課題であったとしても、教科書通りのモデルを作っただけでうまくいくことはほぼないだろう。期待通りに機能させるために必要なロードマップを理解することは、100%自社開発し内製化しない場合でもある程度必要と言える。

無料で学ぶことのできる教材や講座も多くあるため、いきなり難解な専門書で独学しようとして挫折するよりは、やりたいことや興味にあわせて取り組みはじめることが最初の一步ではないだろうか。

以下のセクションにおいてヒントになるようなサイトや資料を掲載したが、主観に基づいて取捨選択したものであり、網羅性に欠けることにご留意いただきたい。また本タスクフォースとして特定の製品やサービスを推奨する意図はないため、各社判断に基づいて適宜ご参照いただきたい。

5.2. (参考) 数理・データサイエンス・AI教育プログラム認定制度

日本企業の競争力強化と持続的成長を実現するためには、データとデジタル技術を活用したビジネスモデルの抜本的な変革（DX：デジタルトランスフォーメーション）が必要であり、それを担う人材を育成する仕組みとして、大学（大学院を除く）、短期大学、高等専

門学校が実施する教育プログラムを認定する「数理・データサイエンス・AI 教育プログラム認定制度」が創設された [14]。

DX を担う人材育成のためのカリキュラムではあるが、何をどのように学んで行くべきかの参考にはなるかもしれない。近い将来、このようなスキルを備えた人材が増えるであろうことを思うと、我々も少なからず知識の底上げが必要であろう。

5.3. 参考文献の紹介

機械学習をテーマとした本はたくさん出版されており、1冊で幅広い知識をカバーするという利点もあるが、より最新の技術や産業応用に関する知見を得るためには文献を探すことになるだろう。

その意味ではこれから AI に取り組もうとする初学者向けの文献を探すことは難しい。代わりに参考になる資料をいくつか挙げてみた。

<機械学習の概要を学ぶ>

総務省ICTスキル総合習得講座 [34]

総務省統計局 高等学校における「情報II」のためのデータサイエンス・データ解析入門 [35]

5.4. 教育プログラムの紹介

独学の場合、体系的に学ぶことのできる教育プログラムの活用は効率的であろう。一方で多種多様なセクションが設定されており、始める前に目標とするゴールや学習すべきコンテンツの取捨選択、期限などを決めておかないと、何を目指していたか、どこまで進んだかを見失う危険性もある。

無料・有料の別は内容の充実度や信頼性には一概に関係があるとは言えない。同じような内容の講座であっても講師によって理解度に差が生じることもあるため、自分に合った講座を選べるよう複数の講師の講座を選択してみるのもよいだろう。

理論学習と同時に、データセットを使って機械学習をさせる実践的な学習も必要である。しかし特に実践的な学習では、解答例がすぐに提示されてしまうような講座では受け身になりがちなので注意が必要である。

5.4.1. UTokyo OpenCourseWare extra (UTokyo OCWx)

東京大学で 2005 年度から始まったオープンコースウェア事業 (UTokyo OCW) で制作された講義コンテンツなどをもとに展開されるオンライン教育コンテンツ。

本講義では、データ分析・データマイニングの基礎について学ぶとともに演習を通して実際にデータを分析するプロセスを知ることができる [36]。

5.4.2. Kaggle

企業や政府などの組織とデータ分析のプロであるデータサイエンティスト/機械学習エンジニアを繋げるプラットフォームで、ユーザーは全世界で 100 万人以上にのぼる。単純なマッチングではなく、企業や政府がコンペ形式で課題を提示し、賞金と引き換えに最も精度の高い分析モデルを買い取るという、クラウドファンディングに近いような仕組みが大きな特徴である。

コンペというと初心者には敷居が高いと思われるが、コンペ時に他のデータサイエンティストが構築したコードが公開されており (Kernel)、コードの書き方に慣れていない初心者でも、Kernel を見ながら実際にコードを書くことで知識を深めることができる。

またプラットフォーム上には有名企業や政府などの組織団体が実際に利用している生データが公開されていることも大きなメリットである。生のデータにはデータ欠損があるため、前処理の段階から実践的なスキルを身に付けることが可能 [37]。

Kaggle の以下の機能は全て無料で利用可能である。

- コンペへの参加
- 「Notebook」：開発不要で機械学習のコードを構築できる
- 「Discussions」：ユーザー同士で質疑応答ができる
- Pythonを使用したコーディング・モデリング
- 「Courses」：Pythonの基礎や機械学習ライブラリの使い方などAI開発に必要な知識が学べる

5.4.3. Papers with code

機械学習の論文、コード、評価表を含む無料でオープンなリソース。

さまざまなタスク (例えば画像分類やテキスト生成など) に対して現時点で最適な性能を発揮する「機械学習モデル」や、スターの多い「コードあり論文」などをランキング形式で紹介している [38]。

<データセット一覧サイト>

各データセットを利用する論文数順で一覧表示される。キーワード検索できるだけでなく、[Modality] (モダリティ: データ種別) / [Task] (タスク: 問題種別) / [Language] (言語) でフィルタリングできる。機械学習の際に「どのデータセットを使えばよいか」に悩む場合に、参考になると思われる [39]。

5.4.4. Python学習教材

5.4.4.1. 東京大学「Pythonプログラミング入門」

東京大学が無料公開している教材で、Python の基本文法からデータの読み込み、機械学習の実装、データの可視化まで幅広い内容を学ぶことができる。多くの練習問題が用意され

ているのが特徴で、自分の理解度を確認しながら進めることができる [40]。

5.4.4.2. 京都大学「プログラミング演習 Python 2021」

京都大学の全学共通科目として実施されるプログラミング演習の教科書として作成された教材で、全 308 ページが無料公開されている。プログラミングの初学者を対象にしており、Python を使ったプログラミングを演習方式で学ぶことができる。Python を使って実際にプログラムを書けるようになることを目的としており、Python の基本操作の習得から Python を使って簡単なプログラムを設計、実装、テストできるところまでカバーしている。教材は図版が豊富で見やすく初学者でも理解しやすいのが特徴 [41]。

5.4.4.3. Google Colaboratory

Google が機械学習の教育および研究用に提供しているインストール不要かつ、すぐに Python や機械学習・深層学習の環境を整えることが出来る無料のサービス。Google アカウントがあれば、簡単に Python の実行環境を構築でき、機械学習に必要な外部ライブラリ (NumPy など) もインストール済みというメリットがある。更に、機械学習など重い処理がハイスピードで行える GPU まで搭載されている [42]。

5.4.5. Coursera (コーセラ)

インターネットを利用した大規模なオンライン講座。Massive Open Online Course の一つ。質の高い講義を無料で受けることが可能。専門知識が少ない人でも内容・説明を理解しやすいよう構成されている。

体系的に学習できるようなコース設計がされているものの、完了には相応の時間を要することが予想される。

<機械学習専門講座> [43]

6. データマネジメントにおけるAI活用の展望

6.1. 臨床開発への応用

これまで臨床試験では、担当者が業務を実施することにより生じる担当者間の品質や判断のばらつき、ノウハウの継承などの問題があった。また、EDC システム以外から入手するデータが複数あり、相互の品質管理は課題となっている。近年ではウェアラブルデバイスや RWD の利活用が増え大容量のデータを処理することも必要になってきている。さまざまな形式のデータや大容量データの処理を可能とする AI を利用することで、より効率的に実行できると考えている (図 6-1-1)。

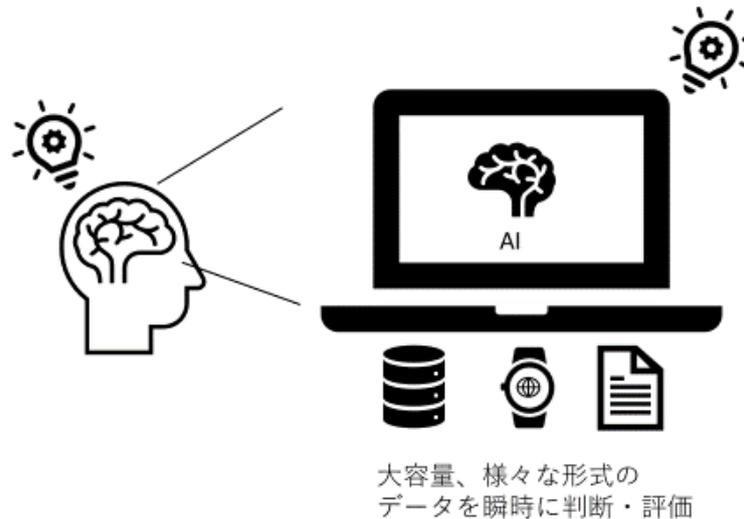


図 6-1-1

既に臨床試験では、広く利用が始まってきている [44] [45] [46]。試験計画段階では、これまでの結果や RWD を用いて試験に最適な対象集団や実施医療機関などを予測することにより、リスクに基づいたアプローチを実行することが出来る。試験開始後は、自然言語処理を用いてカルテ等からデータの抽出や、機械学習を用いたデータクリーニングやデータセットの作成も既に事例が出てきている（表 6-1-1） [47]。

表 6-1-1

試験計画	データマネジメント	解析
<ul style="list-style-type: none"> 対象患者集団の選択 治験薬の投与方法（レジメン）の選択 実施医療機関選定 	<ul style="list-style-type: none"> カルテ等からのデータ抽出 Dataセット作成の自動化（Raw→SDTMの自動化） コーディングの自動化 自動クエリ発行 	<ul style="list-style-type: none"> 解析図表作成の自動化

6.2. 臨床試験におけるAI

ここでは公表されている情報から、近い将来に実際に臨床試験のデータマネジメントに利用できそうな AI の使い方を事例として紹介する [48]。他にもデータマネジメントで利用できそうな AI は数多くあると思われるのであくまでも参考として共有する。これらの事例で使われている AI 技術は決して目新しいものではなく、現在身近に使われている AI 技術を眺めて想像力を働かせれば、データマネジメントで AI を活用できる箇所は少なからず存在するようにも思われた。

6.2.1. eCOA入力タスクに関する通知を最適化するAI

eCOA の入力タスクはタイムリーに実施される必要があり、通知機能が使われることがあるが、過剰な通知は臨床試験参加者の脱落やデータの質の悪化にもつながる恐れがある。

この AI は、チェスロボットが駒の配置を評価して勝つ確率が最大となるように駒を動かすのと同様に、臨床試験参加者に eCOA 入力タスクを最大限に実施させる一方で、通知の総数を最小限にするにはどの通知を臨床試験参加者に送るべきかそしてそれをいつ送るべきかを特定してくれる。この評価のために AI に利用される“記憶された過去の情報”（原文では *state of the system*）は通知時刻、参加者のデータ入力デバイスの GPS 位置、その日にすでに受信した通知の数、および過去の通知への応答などが挙げられる。AI は“記憶された過去の情報”に応じて、動機付けのメッセージまたは期限のリマインダーを送信するか、通知をまったく送信しないこともあり得る。このアプローチは参加者ごとにカスタマイズが可能である。この AI で使われている強化学習は、参加者が eCOA 入力タスクをスケジュールに合わせるのと同時に、役に立たない通知を最小限に抑えることに役立つ。

6.2.2. 画像データの適切性をタイムリーにフィードバックするAI

この AI は画像データの適切性をタイムリーに評価して患者にフィードバックしてくれるため、分散型臨床試験（Decentralized Clinical Trial）において有用である。

分散型臨床試験では、皮膚科の試験であれば患者自身または家族等が撮影した皮膚の画像データを収集することが多い。その際、肌写真が露出オーバーとなっている場合には、AI がフラッシュをオフにするように患者に指示をするため、患者は再撮影することなく迅速に適切な写真撮影を完了することができ、また事後の中央品質管理に要する時間も短縮される。

また神経学的および行動保健学的試験であれば、患者が特定のタスクを実行している様子を患者の家族や介護者等が動画に撮って収集する場合もある。この際に患者の全身がビデオに写っている必要がある場合には、モバイルアプリケーションでビデオが撮影されている間に AI が自動的に確認してくれるため、すぐに必要な調整を行うことができる。

6.2.3. 少数のセンサーで詳細な測定を可能にするAI

時間とともに推移するデータ用に設計された AI は、より少数のセンサーでより詳細な測定を可能にするかもしれない。これらのモデルは、ある身体部分の運動を他の身体部位から収集されたデータを使って推測するように学習させることができ、その結果、全身の運動パターンを評価するために必要なセンサー数を減らすことができるかもしれない。

他のケースでは、単一のセンサーまたは少数のセンサーの組み合わせから AI モデルに学習をさせて、関心のある臨床アウトカムを AI が直接推定することができるようにすることができるかもしれない。例えば、ある研究は、パーキンソン病患者におけるすくみ足は単一のセンサーで検出できることを示唆している。全身に装着される多数のセンサーを単一の

センサーに置き換えることで、運動アウトカムを一般的な生活環境で常に測定できるようになり、これによって稀に発生する兆候等でも測定できるようになる。

6.3. PMS (Post Marketing Surveillance) におけるAI

臨床試験では、CRF のデータレビューやクエリ作成、疾患、薬剤情報のコーディング等のデータマネジメント業務の AI 化に加え、将来的には CRF の情報からデータベースの仕様書の作成、システムバリデーション、およびシステムバリデーション時のテストデータの作成などを AI が担うことも考えられる。PMS のデータマネジメント業務においても、臨床試験と同様 AI の導入は大いに検討可能であると考えられる。上述のデータマネジメント業務に加え、以下の業務においても導入が可能であると考えられる。

6.3.1. PMSの進捗管理

臨床試験では、治験コーディネーターが治験責任医師や分担医師の CRF 作成をサポートするケースが多いが、PMS では一般的に治験コーディネーターのサポートはなく、通常診療を行う医師が CRF 作成からクエリの対応等を行うケースが大半である。

そのため通常医師のサポートは各製薬会社の MR や PMS モニターなどの実施部門が担うことになるが、PMS では SDV は実施できないため、あくまでも医師への調査説明や EDC への入力依頼等の進捗管理、EDC の操作説明など、患者の情報に直接関わらない範囲のサポートに限られる。

各製薬会社の MR や PMS モニターに代わり、今後 AI が PMS の進捗管理を行うことが考えられる。例えば医師の EDC への入力傾向（曜日や時間帯、頻繁に発生するクエリ等）を把握し、各医師に適した注意喚起などを行う、各製薬会社の PMS 担当者が、AI が検出した情報を参考に適切なタイミングで医師のサポートを行う、等の活用方法が考えられる。

6.3.2. ヘルプデスク

PMS では医師、および MR や PMS モニター等の実施部門からの問い合わせに対応するヘルプデスクを各製薬会社で設置するケースが多い。しかし午前 9:00～午後 6:00 の間など、一般的な労働時間内での対応となるケースが一般的である。一方 PMS では通常診療を行う医師が PMS の対応を行うため、ヘルプデスクが利用できない深夜や休日に EDC へのデータ入力等が行われるケースも多くみられる。これらのケースは医師への適切なタイミングでのサポートが行えず、医師への負担増となりうる。また製薬会社としてもタイムリーな情報入手ができない結果となりうる。

しかし近年さまざまな業界で顧客からの問い合わせの一次対応を AI が行うケースが多くなってきた。これらは AI が普段の生活に関わっていることを身近に感じられる一例でもある。PMS においてもヘルプデスクの一次対応を AI が行うような利用方法を検討できれば、PMS のデータ品質の向上や早期の情報入手、および医師の負担低減につながるものと考え

られる。

6.3.3. 有害事象の検知

6.3.3.1. 安全管理部門への報告対象データの検知

薬機法第六十八条の十には「医薬品、医薬部外品、化粧品、医療機器若しくは再生医療等製品の製造販売業者又は外国特例承認取得者は、その製造販売をし、又は第十九条の二、第二十三条の二の十七若しくは第二十三条の三十七の承認を受けた医薬品、医薬部外品、化粧品、医療機器又は再生医療等製品について、当該品目の副作用その他の事由によるものと疑われる疾病、障害又は死亡の発生、当該品目の使用によるものと疑われる感染症の発生その他の医薬品、医薬部外品、化粧品、医療機器又は再生医療等製品の有効性及び安全性に関する事項で厚生労働省令で定めるものを知ったときは、その旨を厚生労働省令で定めるところにより厚生労働大臣に報告しなければならない。」とあり、製造販売業者や外国特例承認取得者は、全従業員に、有害事象の情報を入手した場合 24 時間以内に安全管理部門へ報告することを義務付ける等、各社で報告手順を整備し、法令の遵守に努めている。

近年、PMS においても臨床試験と同様 EDC を使用した調査が主流となり、EDC で電子的に収集されたデータから、プログラムで有害事象情報を検出し、各社の安全性情報管理システムに直接取込ができるデータ形式に自動変換して、安全管理部門に報告するような自動化の流れもできつつある。

しかし EDC とはいえ、入力画面上に担当医師が自由に入力のできるテキストフィールドが設定されるケースも多く、また全例調査や症例数の少ない調査等への対応のため、依然紙の CRF を使用する調査も多く行われている。EDC 上のテキストフィールドに入力されたデータや、紙の CRF に記入された情報から自動的に有害事象情報を検出するのは困難であるため、人の目を介したチェックおよび人の手による安全管理部門への報告が依然行われている。この工程は法令の遵守および各社で設定している安全管理部門への報告期限順守のため、多くのリソースが費やされているのが実情である。

6.3.3.2. 潜在的な有害事象の検知

プログラムで有害事象情報を検出する場合、何を以て有害事象とするか明確な条件の定義づけが必要であり、プログラムで合併症、併用薬剤、臨床検査値等の情報から総合的に判断をして有害事象の検知をすることは困難である。そのため通常はデータマネージャーや安全管理部門の担当者が EDC の情報を入手後にデータレビューを行い、医療機関にクエリを発行して、有害事象の有無を確認する。

この場合、製造販売業者の有害事象情報の入手は、医療機関が上記クエリを確認し、追加の有害事象を入力した後となり、初回の情報入手から安全管理部門の情報入手までどうしても期間が空いてしまう。たとえば AI が EDC 上のさまざまな情報から潜在的な有害事象の可能性を検知し、自動でタイムリーに医療機関にクエリを出すフローを確立することが

できれば、有害事象の報告スピードの大幅な短縮が期待できる。

これらのように“人の目を介したデータレビュー”および“人の手による報告”に、AI を活用できる可能性は多いにあると考える。また、いくつかの AI ベンダーからも、自然言語処理と機械学習によりこれらの作業を AI で行うことは可能であると回答を得ている。有害事象情報の検出や報告の工程に AI を導入する場合に検討すべき内容、および想定される導入フローの例を以下に記載する。

事前検討

少なくとも下記の点について、事前に十分な検討が必要であると考ええる。

<言語>

PMS の業務に AI を導入する場合、AI が日本語に対応していることが前提となる。

<読み込みファイル形式>

EDC から出力した生データを AI で読み込む場合、ファイル形式は SAS データセット、CSV ファイル、Excel ファイル等が想定される。また紙の CRF を読み込む場合は PDF ファイル、Word ファイル等が想定されるが、それらは AI が読み込めるファイル形式であるか確認する必要がある。

また EDC から出力した生データは症例の「背景情報」「併用薬」「有害事象」等パネル毎の出力が想定されるが、紙の CRF は 1 症例/1 ファイルが想定されるため、それぞれ AI が読み込みを行える形式であるか、確認が必要であると考ええる。

<出力ファイル形式>

AI による検出結果がどのような形式で出力されるか、確認が必要であると考ええる。安全性情報管理システムに直接取込ができるデータ形式に変換できると、安全管理部門の作業の効率化にもつながる。

<精度>

有害事象の検出精度が導入を検討している会社のニーズに沿ったものであるかを確認する必要がある。実運用の実績から確認をすることが望ましいが、実運用の実績が無い場合やパイロット運用の実績のみである場合は、限られた情報から判断をすることになるため、十分な事前検討が必要となる。

なお、100%の検出率を AI に求めることは現状では難しいと思われるため、当面は AI の検出結果を人の目でチェックをする工程を想定するのが望ましいと考える。

<学習データの作成>

AI に読み込ませる学習データの確認が必要である。導入を検討している製薬会社側で学習データの作成が必要であれば、どのようなデータを、どのくらいのボリュームで準備をする必要があるかを、事前に AI ベンダーに確認する必要がある。ボリュームに関してはデータベースのレコード数に加え、1 レコードあたりのデータ量（たとえばテキストデータであればどのくらいの長さのテキストデータが 1 つのフィールドに含まれているか、等）も考慮する必要がある。また複数の PMS を横断的に AI がチェックすることが想定される場合は、一連の学習データで横断的なチェックが可能か、あるいは PMS 毎に学習データの作成が必要か、なども確認をする必要がある。

また、AI ベンダーによってはベンダー側で学習データの準備が可能であるケースや、第三者機関から購入するようなケースも考えられるが、少なくとも本項に関連した学習データを製薬会社以外で準備するのは難しいと考えられるため、原則製薬会社側で作成をする必要があると考える。既存のデータベースに格納されているデータがそのまま利用可能か、新たにデータを作成または加工する必要があるか確認をする必要がある。

<AI ベンダーとの契約形態>

製薬会社と AI ベンダーとの直接契約する場合は、AI ベンダーへの Due Diligence、PMS を実施するための契約等の要否の検討が必要であると考えられる。また近年は AI ベンダーと CRO（Contract Research Organization）が提携しているケースもあり、この場合製薬会社は AI ベンダーとの直接契約は行わず、CRO との契約形態も考えられる。いずれのケースにおいても、製薬会社は AI ベンダーの業務状況をどのように把握して管理していくか、事前に検討が必要であると考えられる。

また、AI ベンダーによって、システムにアクセスできるアカウントの発行に対して費用が発生するケース、アカウント数に関係なくシステムの運用期間によって費用が発生するケースなど複数の契約形態を準備している場合もあるため、どのような契約形態が最適か十分な検討が必要である。

<Audit への対応>

PMS の業務に AI を導入する場合、アウトプットの妥当性について第三者への説明が必要になるケースが考えられる。AI 導入時のシステムバリデーション、役割分担、作業工程、結果などについて、第三者への説明が可能であるか事前に検討が必要であると考えられる。

<AI のシステムバリデーションおよび教育記録>

AI をシステムの位置づけとして考える場合、システムバリデーションの観点で人間がどのように AI の検証を行い、記録を残すか検討を行う必要がある。ただし、AI をシステムとしてではなく人間の代理として考える場合、システムバリデーション記録ではなく人間と同様トレーニングの記録を提示することも考えられる。また、いくつかの AI ベンダーから、

ある程度の AI の情報検出メカニズムを、人間が判読可能な形でアウトプットすることも可能であると回答を得ており、これらの資料も利用が可能であると考えられる。

<導入、運用費用および期間>

導入、運用にかかる費用、導入にかかる期間に加え、AI 導入後も製薬会社側に AI を管理するための一定のリソースが必要となることが想定されるため、事前に検討が必要であるとする。

<データマネジメントのみへの導入か>

AI はデータマネジメントに特化した仕組みではなく、カスタマーサービス、マーケティングなどさまざまな分野での活用が期待できる。最大限の導入効果を得るためには、部門横断的な検討を行うことも必要であるとする。

<導入フロー例>

1. 業務内容の検討、契約
2. 学習データの準備

AI ベンダー側、もしくは製薬会社側で学習データの準備を行う。

3. AI の設定

AI ベンダーはテスト環境において AI の設定を行い、システムバリデーションを行う。

4. AI のトレーニングを行う。

学習データを使用した AI のトレーニングを行い、検出結果の精度を高める。

5. 実装

AI ベンダーは本番環境への実装を行う。

6. メンテナンス

必要に応じて、定期的な AI のトレーニング等を行う。

6.3.3.3. これからの市販後安全対策

ビッグデータの時代となり、さまざまな業種でデータ活用が進む中、平成 30 年 4 月 1 日に施行された改正 GPSP 省令では、新たに「製造販売後データベース調査」の定義を設けるなど、市販後の安全性情報の収集においてもビッグデータの活用は着実に進んできていると言える。また新型コロナウイルス感染症の影響で、PMS や自発報告等の手法による迅速な情報収集が困難となっている中、令和 4 年 3 月に PMDA において「SNS を医薬品安全対策の一環として活用するための調査業務調達仕様書」が作成され、SNS の利活用拡大が見込まれる中、これらの情報を市販後安全対策の一環として活用するための準備が進められている。

これらのような日々蓄積され変化し、形式もさまざまにかつ膨大なデータ量を人の手で解析し、評価を行うのは非現実的であり、AI を用いて効率よく解析を行い、医薬関係者等にタイムリーに情報提供を行うことで、市販後安全対策の確保を図ることが、今後求められてくると考えられる。

おわりに

2022 年 11 月に登場した OpenAI の ChatGPT は世界に大きな衝撃を与えた。プログラミング、音楽、小説、脚本、詩、歌詞や作文などの創作活動等が行える他、特定のテストに対しては人間と同水準かそれ以上の回答が行えるなど幅広い機能を持ち合わせているこのジェネレーティブ AI はチャットボットを遥かに超える存在に進化してしまった。更に Microsoft は本年 3 月に Word、PowerPoint、Outlook、Teams などを含むすべての Microsoft 365 アプリに ChatGPT を統合させた Microsoft 365 Copilot を発表した。今後ジェネレーティブ AI はごく普通に自然にビジネスや日常生活で使われていくことになるだろう。

2022 年度のタスクフォースとして約 1 年間、データマネジメント領域での AI 利用のハードルを下げるための成果物作りに取り組んできた。具体的なイメージを持っていただくためにユースケースを紹介することはタスクフォース当初から考えてきたことで、紙面の都合や著作権等の理由で引用または転載できないものも多々あるため、原著が参照できるものについては原著を参照していただきたい。

この成果物作成は、タスクフォースメンバーにとっても、あらためて技術的な観点、規制や倫理的な観点、ビジネス的な観点等から AI を見つめる良い経験になった。本成果物の読者の皆さまにも是非同じようなエクスペリエンスをしていただければと考えている。

タスクフォースに Python のプログラムが書けるメンバーがいたこと、PMS における安全性情報報告で苦勞しているメンバーがいたことから、Appendix 2 のプロトタイピングの構想が湧いてきて、実施してみたところ思っていた以上の成果を上げることができた。是非ご自分の仕事でも AI を活用できる課題を特定してプロトタイピングされることをお勧めする。

今後 AI 社会の中で生きていくためには、AI リテラシーを上げていかなければならず、また課題解決能力を高めていかなければならない。自分たちの仕事にどのような AI 技術がどのように活用できるのか想像力を働かせて考えていく必要がある。ChatGPT が一見自然に見えるが事実とは異なる回答を生成することが問題視されているように、我々も業務上 AI を利用するにはそこに内在するリスクを認識して、十分な対策を講じておかなければならない。我々の扱っているデータは医薬品のリスクベネフィットの判断のもとになるクリティカルな情報であることを忘れてはならない。

本書を読んでくださった方々が、大いに前向きにしかし前のめりにはならず、AI 時代のデータマネジメントを切り拓いていってくださることを、タスクフォースメンバー一同

心から願っている。

Appendix 1: 緒論

近年、コンピューターの性能が高くなるにつれ、一度に膨大な量のデータを分析したり、複雑な処理を手軽にできるようになってきた。それに伴い、機械学習や深層学習の技術も進み、同時に AI が進歩してきた [49]。この章では AI の技術用語について「AI ってなに？」の記載内容をおさらいするとともに臨床試験への応用を概説する。

A1.1. 機械学習

機械学習はプログラミングすることなく、データからモデルを学習させることであり、教師データあり学習、教師データなし学習および強化学習に分類できる。

【教師あり学習 (Supervised Learning)】

教師あり学習は入力値と正しい出力値のペアデータ(ラベル付きデータ)を使用し、入力された値に対して適切な値が出力されるモデルを学習する方法

例：分類、回帰

【教師なし学習(Unsupervised Learning)】

「教師」としてのラベル付きデータを用いない手法である。教師なし学習では、ラベル付きのデータを用いる代わりに、データのパターンを調べる。

例：クラスタリングや次元削除

【強化学習(Reinforcement Learning)】

教師あり学習のように過去の観測に対して正解が与えられるものではなく、どれだけ良いかの指標のみが与えられる。

1950 年代に Arthur Samuel によってチェッカープログラム (Samuel Checkers-playing Program) が開発され [50]、このとき以降、『機械学習 (Machine Learning)』という言葉が使われるようになった。その後今日までに機械学習は、発展を遂げ、さまざまな手法が構築されている。以降の章では、その中でも代表的な手法について簡単に紹介する。

A1.1.1. 教師あり学習の手法

A1.1.1.1. 決定木 (decision tree)

決定木は、分類木と回帰木を組み合わせた教師ありの学習方法である。分類木は、簡単な質問に対する Yes/No といったカテゴリカルな回答となるものを指し、回帰木は、数値のような連続値が回答となるものを指している。つまり、カテゴリカルデータおよび連続値、両方とも設定することが可能である。

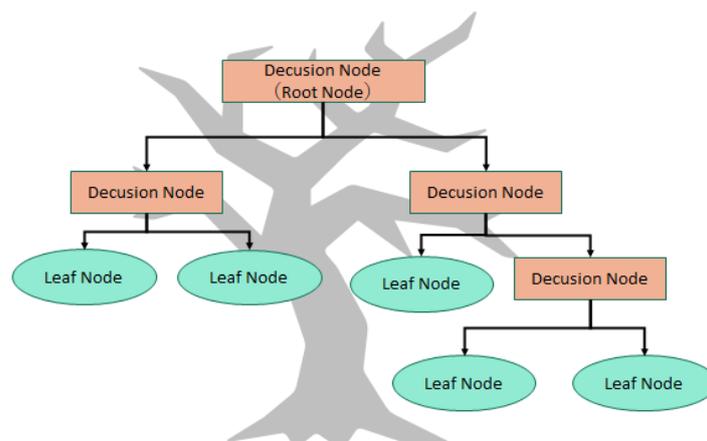


図 A1-1-1

決定木は、3種類のノードで説明される。ルートノードは、最初のデータセットを指し、データは2つのサブツリーに分割される。Decision ノードは、質問を指しその回答によって、次の質問に進むかが決定される。Leaf ノードは、最終的な結果であり、これ以上は、分岐されることはない（図 A1-1-1） [51]。

決定木学習では、出来るだけ詳細に分割するために、このノード分割が繰り返し実施されるが、決定木内の枝が多くなってしまうと、異常なデータまでも学習してしまう。そのため、枝刈り（Pruning）と呼ばれる不要なノードを削除する作業を行う必要がある [52]。

決定木の手法を更に進化させたものがランダムフォレスト [53]といわれる手法である。ランダムフォレストは、決定木を複数組み合わせたもので、訓練データを複数に分けそれぞれで決定木を作成しそれらを組み合わせることにより、回答の精度を上げることを可能としている [49]。

A1.1.1.2. ロジスティック回帰

ロジスティック回帰は、カテゴリカルな結果を予測する分類手法である。ロジスティック回帰の結果は、0 から 1 の間の結果となる。得られた結果が閾値（例えば 0.5）より大きければ、1 となり、閾値より小さければ 0 となる（図 A1-1-2） [54]。

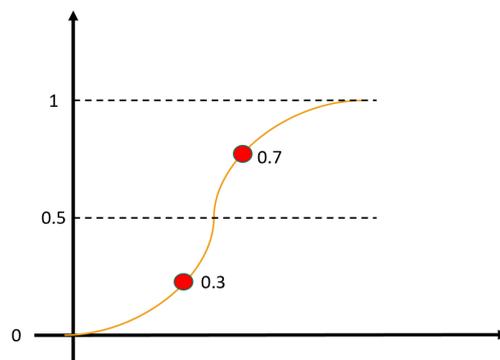


図 A1-1-2

A1.1.1.3. K-近傍法 (K-Nearest Neighbor)

K-近傍法は、ノンパラメトリックな教師あり学習の手法である。例えば、A という分類と B という分類の 2 つに分類できるデータがあり、新たに得られたデータは、A と B のどちらに分類されるのかを調べたいとする。

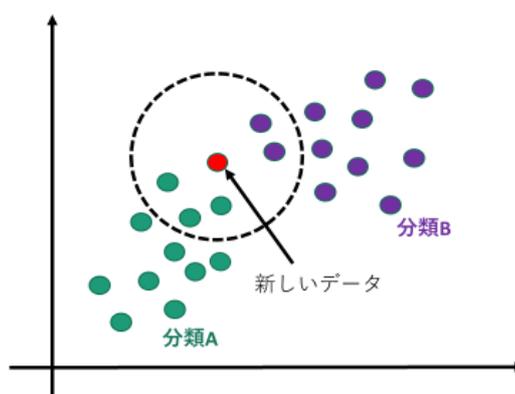


図 A1-1-3

まず、近傍の数を指定し (例 $k=5$ 個)、データ間のユークリッド距離を算出する。距離の近いものから上位 5 個を見てみると、A に分類されているものが 3 個、B に分類されているものが 2 個となり、新たに得られたデータは、A に分類されると判断できる (図 A1-1-3)。

この手法の欠点としては、 k の値の設定により、結果が異なることがあり、更に学習データが少ないと精度が下がる。また、すべてのデータのユークリッド距離を算出するためマシンへの負荷がかかることが知られている [55] [52]。

A1.1.1.4. naïve ベイズ (単純ベイズ)

naïve ベイズは、ベイズの定理に基づく確率から推定する手法である。事前の知識で仮説の確率を決定するが、既知の情報が独立していると仮定しているため、naïve（単純）と言われる [56]。

与えられた学習データから、すべての確立推定値を事前に計算して、保存できるため、高速予測が可能である [52]。

A1.1.1.5. 人工ニューラルネットワーク

人工ニューラルネットワークは、ニューロンをモデルとした、非線形分離を可能とする手法である [49]。

ニューロンの樹状突起 (Dendrite) を Input、細胞核 (Cell Nucleus) をノード、シナプス (Synapse) を重みづけ、軸索 (Axon) を Output で表している。

人工ニューラルネットワークは、3つの層に分かれている (図 A1-1-4)。

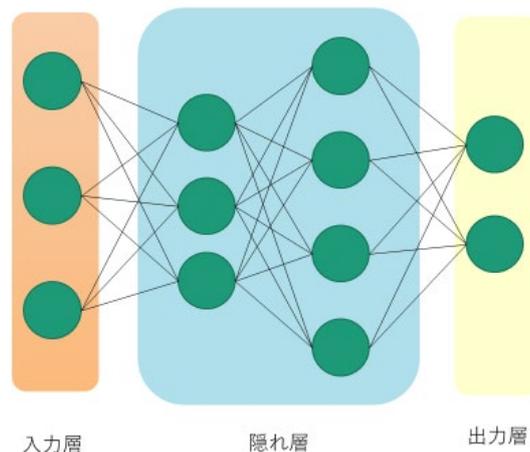


図 A1-1-4

入力層は、データの入力であり、出力層は、結果の出力となる。中間にある隠れ層で分類のパターンを探す層となる [57] [52]。

深層学習では、いくつもの人工ニューラルネットワークを組み合わせた多層ニューラルネットワークを用いている [49]。

人工ニューラルネットワークは、並列処理能力に優れているが、適切なネットワーク構造が作成できるかは、作成者の経験や能力によって大きく変わってしまうことが知られている [57]。

A1.1.1.6. Support Vector Machine (SVM)

データを分類する際、分類線は無数に存在する場合がある (図 A1-1-5 水色線)。この分類線の内、2つの分類の真ん中にある線を超平面と呼び (図 A1-1-5 赤直線)、誤分類が最

も少ない分類線となる。超平面から一番近いデータポイントをサポートベクターという。サポートベクターから超平面の距離の和 (γ) が最大となるような超平面を算出する。

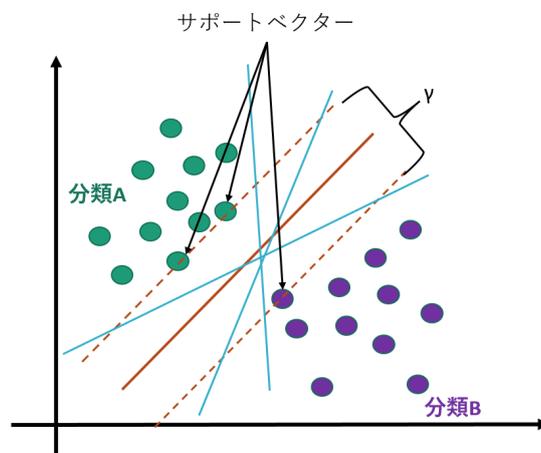


図 A1-1-5

このような超平面モデルを元に知りたいデータがどちらに分類されるのか、推定することが出来る [58] [52]。

A1.1.2. モデルについて

モデルの構築については、多数の手法が考案されており、Python ライブラリ等の様々なツールを活用することで初学者でもモデルを構築することができる。ゼロからモデルを構築することは、費用、人的資源及び時間といった面で困難である場合は、学習済みモデルを利用し微調整を行う方法などもある。いずれの場合も、適切に意図したモデルが構築できているかの確認や、本質的に理解するためには数学や統計の知識が欠かせず、専門家の助力も必要と考えられる。

A1.1.3. モデルの構築と評価

モデルを作成するために全てのラベル付きデータを学習に使用した場合、未知のデータを用いて得られた予測結果が真に正しいか否かを判定するには、得られた結果に対してすべて目視確認することが必要となる。そのため、一般的にはラベル付きデータを①学習フェーズに用いる学習データ、②学習済みモデルの初期評価およびハイパーパラメーターのチューニングのために用いる検証データ、そして③学習済みモデルの予測結果を評価するテストデータの3つに分割し、モデルの性能を評価する。また、モデルの評価にはさまざまな指標が利用でき、目的に合った指標を用いる必要がある。例えば Appendix2 で実施している医学事象検出においては「誤って検出されること（偽陽性）」よりも「検出漏れ（偽陰性）」をどれだけ低くできるかを重視する必要がある、偽陰性の割合を確認できる再現率 (Recall)

が重要な評価指標となる。

A1.2. 自然言語処理

自然言語とは人間が日常的に読み書きや会話中に使用している言語である。自然言語で記載されたデータをコンピューターが正しく扱えるように処理することを自然言語処理と呼ぶ。自然言語処理を行うことで、作業の高速化や人間の作業による見落としを回避することができ、業務をより効率化することができる。臨床試験では EDC の導入により、フリーテキストのデータを収集する機会は少なくなってきたが、医療機関から回答されたクエリ回答に対して自然言語処理を行うことで再クエリ発行の可否を判断するサポートになるであろう。また、紙 CRF を使用している試験では医師が記載したフリーテキストの中から有害事象を示唆する文章や用語を自然言語処理を使用してピックアップすることが可能となり、安全性部門が集める安全性データから特定の疾患の傾向を示すシグナルを検出することも可能かもしれない。

以降の章では、自然言語処理で使われる以下のステップを説明する。

- 形態素解析
- 文章¹¹のベクトル化
- 分類処理

A1.2.1. データの前処理

学習用に準備したデータに対して、必要に応じて、データの前処理などを行う必要がある。データを扱いやすいように変形したり、複数のファイルに分かれているデータを 1 つにまとめたり、欠測値の保管や単位を揃えるなどの対応 [16]が必要になる。また、ラベル付けが必要な場合はラベル付けを行うなどの作業も必要になる。

非構造化データであるテキストデータ（文章）の場合は、まず単語に切り分ける必要がある。ブログや SNS などをソースとする場合はクリーニングでテキスト中の html のタグなど文章に直接関係のないデータの除去が必要になる。「形態素解析器」と呼ばれるライブラリを用いることで単語の切り分けと同時に品詞情報の付加、単語の正規化で行う活用形を原型に変換することができる [16]。そして接続詞、助詞などの学習に役立たない単語（ストップワード）を形態素解析で付加された品詞情報を用いて識別、除去する。単語をベクトルで表現（変換）することで統計処理可能となる。

¹¹ 一般的には「文書」のベクトル化が行われることが多いが、本成果物においては「文章」を対象としてベクトル化と分類処理を実施しているため、「文章」という表現で統一している。

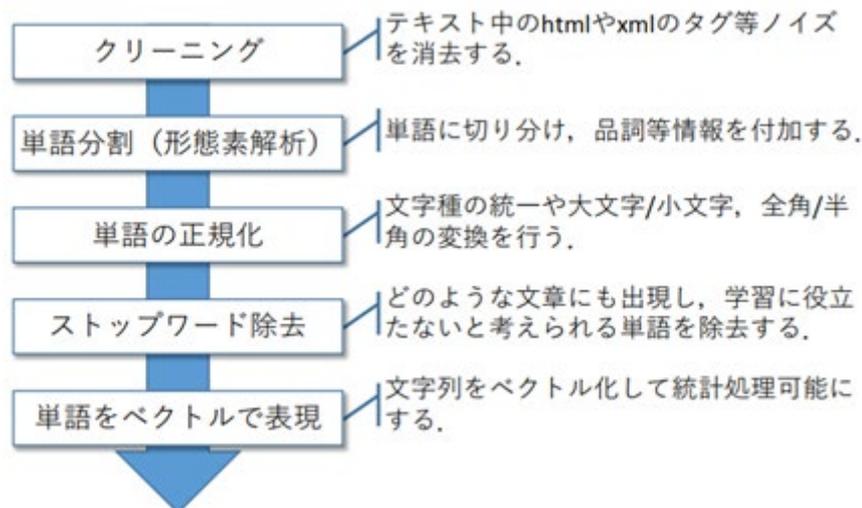


図 A1.2.1 テキストデータの前処理の概要 [16]

A1.2.2. 形態素解析

文章に対し、自然言語処理を実行する際、文章内にどのような単語がどのような頻度で出現しているかを知ることが重要である。形態素解析とは、文章を「形態素」と呼ばれるレベルに分解し、各形態素に対して品詞を付与する処理である。英語は単語毎にスペースで区切られているが、日本語は単語が並び、文章を構成しており、文章単位で処理することができないために、形態素解析が特に必要である。Python では MeCab や Janome といった形態素解析器が使用できる。

例) 文章： 2022 年 11 月 1 日、朝から発熱あり¹²

Janome による形態素解析結果：

2022	名詞,数,*,2022,*,*
年	名詞,接尾,助数詞,年,ネン,ネン
11	名詞,数,*,11,*,*
月	名詞,一般,*,月,ツキ,ツキ
1	名詞,数,*,1,*,*
日	名詞,接尾,助数詞,日,ニチ,ニチ
,	記号,読点,*,、,、,、
朝	名詞,副詞可能,*,朝,アサ,アサ
から	助詞,格助詞,一般,から,カラ,カラ
発熱	名詞,サ変接続,*,発熱,ハツネツ,ハツネツ
あり	動詞,自立,*,ある,アリ,アリ

¹² Google Colab Link: [JPMA_2022_TF_1_1_demo_形態素解析.ipynb](https://colab.research.google.com/drive/1JpMA_2022_TF_1_1_demo_形態素解析.ipynb)

MeCab (IPAdic-NEologd-JMeDic) による形態素解析結果 :

2022 年 名詞,固有名詞,一般,*,*,*,2022年,ニセンニジュウニネン,ニセンニジュウニネン
 11 月 1 日 名詞,固有名詞,一般,*,*,*,11 月 1 日,ジュウイチガツツイタチ,ジューイチガツツ
 イタチ
 、 記号,読点,*,*,*,、,、,、
 朝 名詞,副詞可能,*,*,*,朝,アサ,アサ
 から 助詞,格助詞,一般,*,*,*,から,カラ,カラ
 発熱 名詞,サ変名詞,*,*,*,はつねつ;icd=R509;lv=S/freq=高頻度;発熱,はつねつ,は
 つねつ,1267628
 あり 動詞,自立,*,*,五段・ラ行,連用形,ある,アリ,アリ

A1.2.3. 文章のベクトル化

文章を形態素に分割後、そのままでは解析に用いることが難しいため、各形態素に対して数学的な処理を可能とするためにベクトル化を行う。文章をベクトル化するための代表的な手法として、Python では Bag of words、TF-IDF、word2vec などが使用できる。

A1.2.3.1. Bag of words と TF-IDF

Bag of words は単語の出現のみを考慮して文章をベクトル表現する手法であり、TF-IDF は、単語の出現頻度を考慮して文章をベクトル表現する手法である。TF-IDF において TF (Term Frequency) は文章中の単語の出現頻度を表し、IDF (Inverse Document Frequency) はその単語がどれだけ特定の文章に偏って出現しているかを示している。これは、高頻度および、出現傾向に偏りのある単語が重要であるという人間の直感を表現した手法である。ただし、数値や助詞なども全ての単語が対象となるので、誤って類似していると判定されてしまうケースがあることや、記載の揺れで類似していないと判断されてしまうこともあるため工夫が必要である。なお、いずれの手法も単語の出現する順番は考慮されない [59] [60]。

A1.2.3.2. Word 2Vec

Word2Vec は、周辺単語を入力して次単語を予測する形の学習である。単語の分散表現の形で表現するため、単語の類似性を考慮することが出来る。

つまり、表記が異なっても、意味が同じであれば、似た分散表現として結果を得ることが出来る。したがって、表現が多様な文章の場合に向いている [59] [60]。

A1.2.4. BERT (Bidirectional Encoder Representations from Transformers)

BERT は 2018 年に Google から発表されたモデルである。BERT のモデルは事前学習済みモデルと fine-tuning の 2 段階にわけることができる。前述した word2vec では同じ単語は文脈の意味に関係なく、同じベクトルに変換されるが、BERT は文脈に依存しベクトル

変換されるため、同じ単語でも文脈により別のベクトルに変換される。

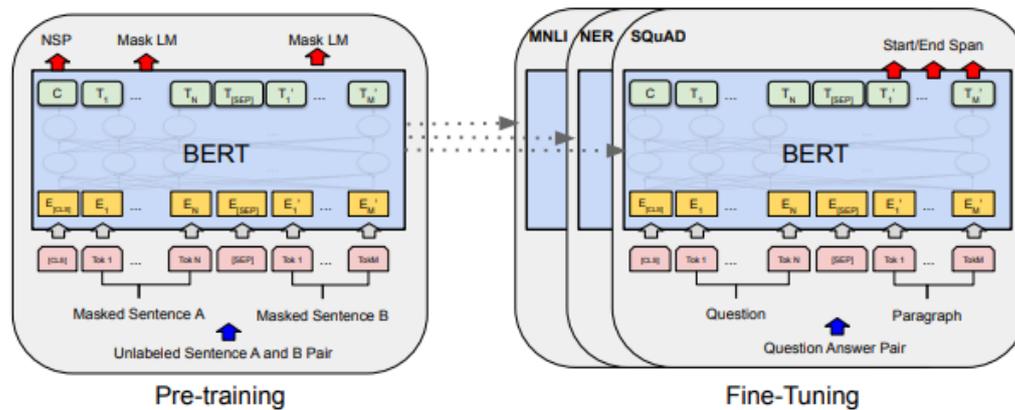


図 A2-1-4-1 [61]

A1.2.5. 固有表現抽出

地名、人名、組織名などの固有名詞に、日付、時刻、数量などを加えたものを固有表現という。テキスト内の固有表現を素性ごとにラベリングをしておくことで、自然言語から情報を検索・抽出することが可能である。

英語の場合は、地名、人名、組織名などの先頭が大文字になっているため、日本語より高い精度で抽出が可能である。日本語においては、ラベル付けする前後 5 単語について分析することで、高い精度でのラベリングを可能としている [59] [60]。

Appendix 2: 自然言語処理技術を活用した有害事象判定モデルの作成

「1 データマネジメント領域におけるユースケース」で紹介したユースケースのように、データマネジメントの業務においても AI の導入が進み始めている。しかしながら、これらは英数字を対象としたものであり、日本語を対象とした AI 活用事例は見つけることができなかった。本邦においては治験のデータ収集は英語を用いることが一般的となっている一方で、PMS のデータ収集は日本語を用いることが一般的となっている。「6.3 PMS (Post Marketing Surveillance) における AI」で述べたように、PMS では検知した有害事象を報告する義務があるため、フリーテキストで記載されている内容をすべて人の目で確認しているのが現状であり、適切に検出するためにはリソースのみならず確認者の練度も求められている。そこで本タスクフォースでは DM 業務における日本語 AI のトライアルとして、有害事象と判定されるべき記載が AI によってどの程度の精度で検出が可能であるか、プロトタイプモデルを作成して検討した。本章では、実際のプログラムの構成と判定結果を紹介する。なお、プログラムはすべて Colabatory [42] 上で作成し、GitHub に公開¹³している。Google アカウントを取得すれば誰でも無料で実行可能¹⁴であるため、実際の挙動や検討すべき事項の確認など、今後の AI 活用を検討する際の参考となれば幸いである¹⁵。

A2.1. プログラムの作成と実行

本タスクフォースでは特定の文章が「有害事象あり」と「有害事象なし」のいずれに分類されるかを判定する、文章分類タスクをプロトタイプの対象とした。この文章分類を行うプログラムとして、2種類のプログラムを作成している。1つは機械学習を使用せず、単語の出現頻度から文章の類似度を計算するコサイン類似度を用いるプログラムであり、これは「有害事象あり」のグループでは有害事象に特徴的な用語を含む文章のベクトル表現は類似しているであろうという仮定に基づいている。もう1つは機械学習モデルを組み合わせたアンサンブル手法を用いるプログラムである。

A2.1.1. データの用意

対象とする文章は、主に厚生科学審議会（予防接種・ワクチン分科会 副反応検討部会）の資料 [62] から取得し、特定のワクチン名や用語に影響されないように加工を行った。また、新たにタスクフォースで用意した文章も追加し、約 5000 件の文章を用意した。これらの文章に 2 種類のラベル付けを行っている。LABEL1 では、その文章のみで何らかの事象の悪化や確認すべき事象があると判断したものを 1:Positive とし、LABEL2 はその中でも更に重篤性の高い事象や妊娠等の注目すべき事象が含まれるものを 1:Positive とした。それぞ

¹³ GitHub Repository: [GitHub - Takumi173/JPMA2022TF1-1](https://github.com/Takumi173/JPMA2022TF1-1)

¹⁴ プログラム間でデータ受け渡す目的で Google drive の接続を必要とします。必要に応じてプログラム実行用の Google アカウントを取得してください。

¹⁵ プログラムの作者および製薬協は、本プログラムの実行によって生じたいかなる損害に対して、一切の責任を負いません。

れの件数を表 A2-1-1-1 に示す¹⁶。

表 A2-1-1-1

	LABEL1	LABEL2
1:Positive	2535 件	635 件
0:Negative	2484 件	4384 件

LABEL2 ではその特性上 1:Positive となるデータが少なく、データの偏りを緩和するために LABEL1 が Positive であるデータのみを使用して学習を行うこととした。LABEL2 の学習において実際に使用した件数は表 A2-1-1-2 の通りである。

表 A2-1-1-2

	LABEL2
1:Positive	635 件
0:Negative	1900 件

なお、1:Positive と判断する基準は各組織や個人によって揺らぎがあると思われ、本タスクフォースで用意したラベル付きデータは、各組織や個人の判断基準に照らし合わせて必ずしも適切とは言えないことに注意されたい。機械学習では付与したラベルをもとにモデルが構築されるため、付与したラベルの適切性をユーザー側で確認しておくことが必要となる。

A2.1.2. 文章のベクトル化

A2.1.2.1. 表記の統一

各文章の形態素解析を行う前に、文字種や表記揺れを統一するために以下のような処理を行った。

1. `jaconv` ライブラリを使用した半角・全角等の文字の正規化
2. `unicodedata` ライブラリを使用したユニコード文字の正規化
3. `neologdn` ライブラリを使用した反復表記などの不要な表現の除去や、ハイフンやダッシュ、チルダ等の記号文字の正規化

また、今回のプロトタイピングでは処理を容易にするため、数字はすべてゼロに置換している。フリーテキストの検出においては臨床検査値の高低などを読み取って判断することも必要になる場合があるが、今回のプロトタイピングにおいては処理を容易にするために値の高低による判定は対象外とした。

¹⁶ 生データは GitHub からダウンロード可能：<https://github.com/Takumi173/JPMA2022TF1-1/releases/download/20221226/Training.txt>

A2.1.2.2. 形態素解析

形態素解析器¹⁷は MeCab を使用し、システム辞書として mecab-ipadic-NEologd [63] を使用、ユーザー辞書として万病辞書 [64] を使用した。形態素解析後、形態素に原形が定義されている場合は原形に置き換えを行い、表現を統一した。不要な情報を落とすために品詞による単語の絞り込みを行い、名詞、形容詞、助動詞、動詞の 4 つの品詞を意義のある品詞として残した。

A2.1.2.3. ベクトル化

文章の形態素解析を行った後、Bag of Words と TF-IDF によるベクトル化を行った。また、東北大学の Wikipedia Entity Vectors [65] を参照した分散表現も取得し、この分散表現と TF-IDF を掛け合わせたベクトルも作成した¹⁸。

¹⁷ 形態素解析については「A1.2.2 形態素解析」を参照。

¹⁸ Google Colab Link: [JPMA 2022 TF 1.1 demo \(1\)前処理からベクトル化まで.ipynb](#)

A2.1.3. 判定モデルの構築

A2.1.3.1. コサイン類似度を用いて判定するモデル

TF-IDF 値を用いてコサイン類似度を計算し、予測したい文章が LABEL1 の Positive ラベル群と Negative ラベル群のいずれに近いかを判定するプログラムを作成した¹⁹ (図 A2-1-3-1-1)。

プログラムの予測手順は以下の通り：

1. ラベル付き文章を参照データとして予測したい文章を組み合わせる各文章の TF-IDF 値を算出する
2. 予測したい1つの文章に対して、総当たりでコサイン類似度の計算を行う
3. 最も大きなコサイン類似度を示した文章のラベルを予測結果とする

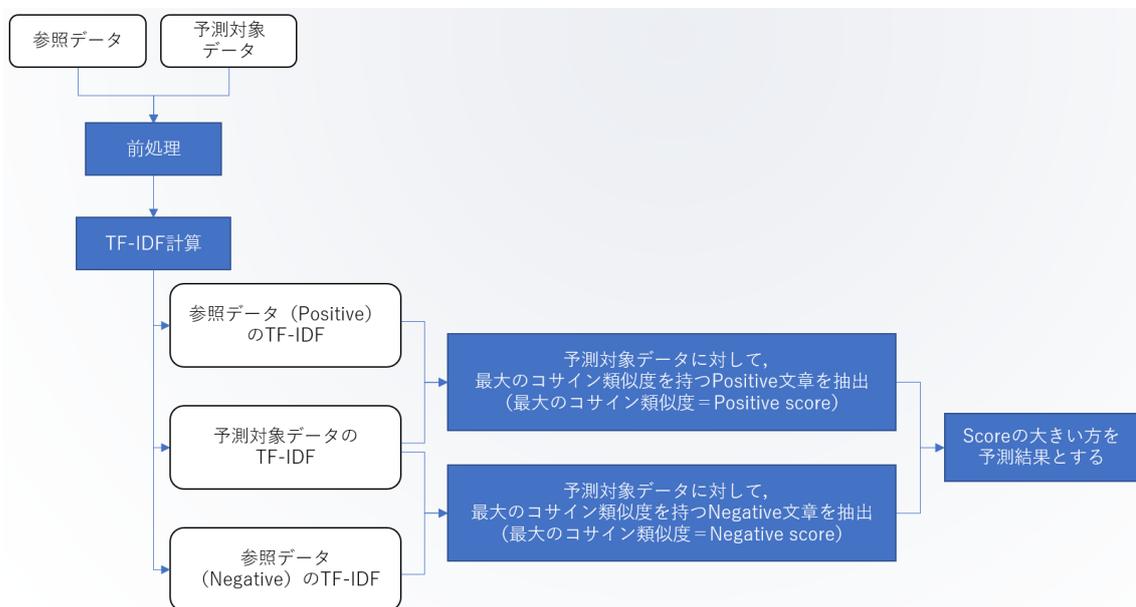


図 A2-1-3-1-1

A2.1.3.2. アンサンブルモデル

「A2.1.2 文章のベクトル化」で作成した4つのベクトル (Bag of Words、TF-IDF、分散表現、分散表現×TF-IDF) に対し、10種類 (ロジスティック回帰、リッジ回帰、線形 SVM、SVM、ナイーブベイズ、k 近傍法、ランダムフォレスト、Ada Boost、Light GBM、ニューラルネットワーク) の手法を用いて、F1 スコアを評価指標²⁰にモデルを構築した²¹。また、ベクトル化前の前処理済み文章を使用し、東北大学の事前学習済み日本語 BERT モデル

¹⁹ Google Colab Link: [JPMA 2022 TF 1 1 demo \(a\)コサイン類似度による比較.ipynb](#)

²⁰ Recall ではなく F1 スコアを評価指標としているのは、Recall は偽陽性の割合を全く考慮しないため、極端な例では全データに対して Positive 判定することが最良の Recall となるためである。

²¹ Google Colab Link: [JPMA 2022 TF 1 1 demo \(2\)Grid Search と結果の確認まで.ipynb](#)

[66]を使用して Fine-tuning による分類モデルの構築を行った²²。その後のアンサンブルモデルの構築においては、4つのベクトルに対して最良のスコアとなるモデルを手法が重複しないように選択した。BERTモデルの Fine-tuning では 5-fold cross validation を実施した中で最良の結果が得られたモデルをアンサンブルモデルに追加した²³ (図 A2-1-3-2-1)。

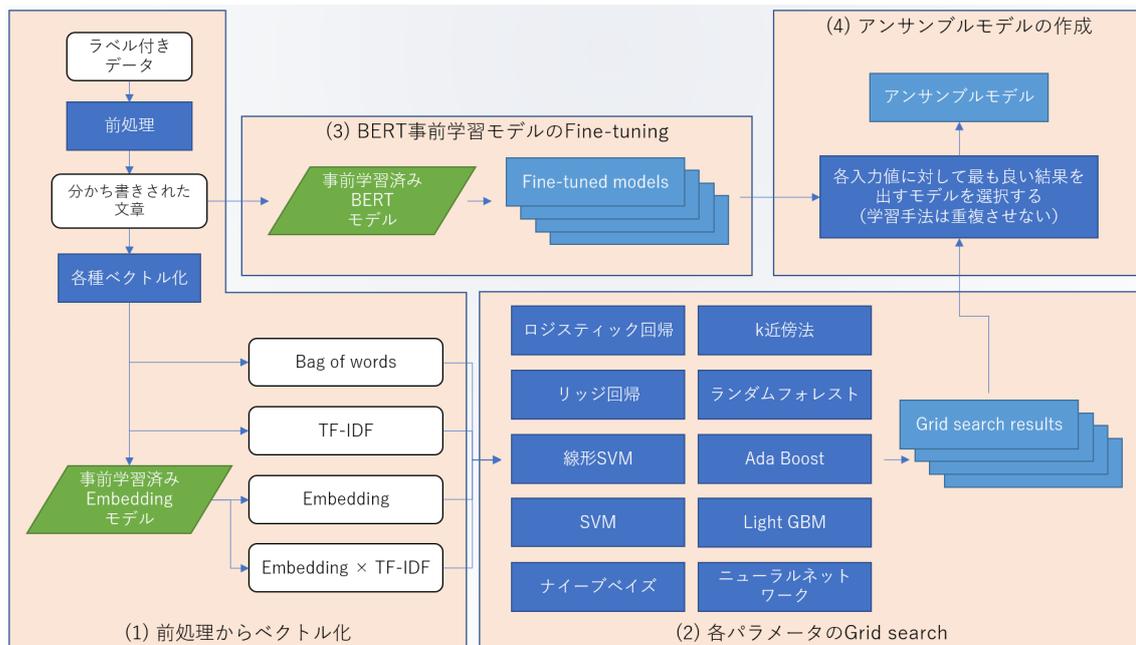


図 A2-1-3-2-1

予測時には、予測したい文章に対してモデルの構築時と同じ手順でベクトルを取得し、選択したそれぞれのモデルから得られた結果を統合してアンサンブルモデルの結果とした。なお、Bag of Words と TF-IDF のモデルにおいては、訓練データに存在しない未知の単語は評価できないため、予測前に除外している。また、TF-IDF 値は全文章中の頻度を計算するため、訓練データと予測対象データを組み合わせて再計算した値を入力値としている。アンサンブルモデルの作成において、今回対象としているのはフリーテキスト欄に記載された有害事象の検出であり、判定の精度向上よりも「報告漏れ」に通じる偽陰性を防ぐことが良いと考え、いずれかのモデルが Positive と判定した場合は他のモデルが Negative と判定していても総合的に Positive 判定をする仕組みとした。また、判定には影響しないが、それぞれのモデルから予測スコアとなる数値を取得し、その総合値をアンサンブルモデルのスコアとして表示できるようにした。加えて、判定の内容を説明する仕組みとして、アンサンブルモデルがどの単語に注目して判定を行っているのか可視化する LIME [67]も実装し

²² Google Colab Link: [JPMA 2022 TF 1 1 demo \(3\)BERT 分類モデル.ipynb](#)

²³ Google Colab Link: [JPMA 2022 TF 1 1 demo \(4\)アンサンブルモデルの作成と検証.ipynb](#)

た²⁴ (図 A2-1-3-2-2)。

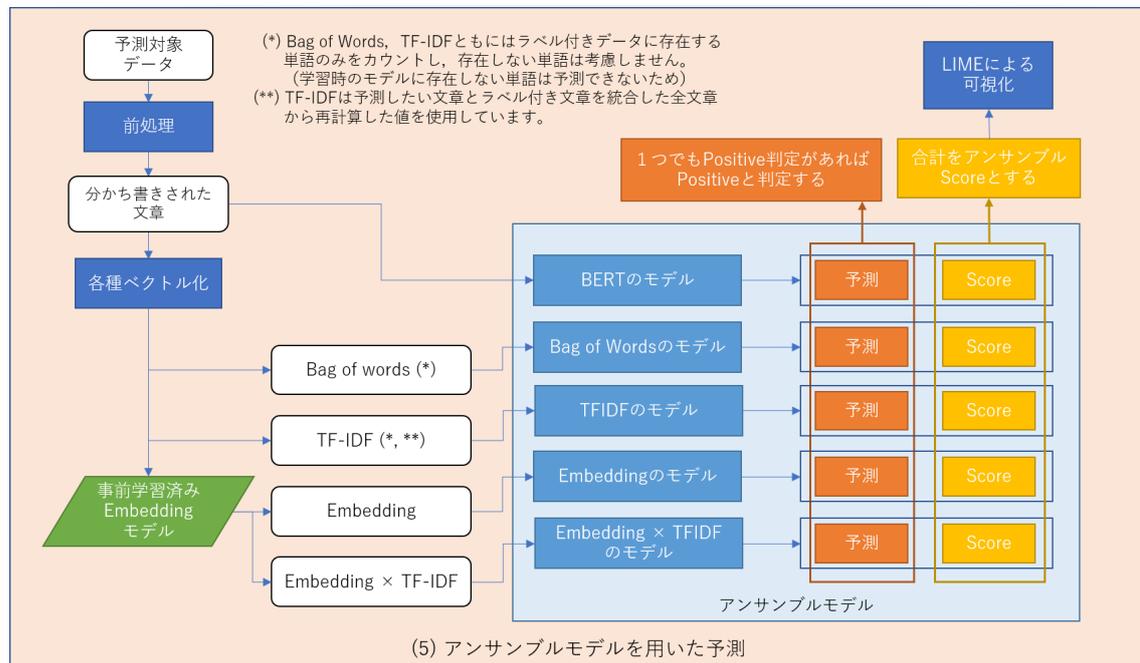


図 A2-1-3-2-2

²⁴ Google Colab Link: [JPMA 2022 TF 1 1 demo \(5\)アンサンブルモデルを用いた予測.ipynb](#)
 予測を行うにはアンサンブルモデルを構築するまでの(1)~(4)のプログラム実行が必要であるが、手早く確認してみたい方向けに、タスクフォースで用意した学習結果を利用してアンサンブルモデルを試せるプログラムも用意した。
 Google Colab Link: [JPMA 2022 TF 1 1 demo \(5+\)アンサンブルモデルを用いた予測 \(事前学習済みデータを使用\) .ipynb](#)

A2.1.4. 評価結果

コサイン類似度を用いて判定するモデルではラベル付きデータの 85%を参照データ、残りの 15%を予測対象データとして評価を行った。アンサンブルモデルではラベル付きデータの 85%を学習&検証データとし、残りの 15%をテストデータとして評価を行った(図 A2-1-4-1)。

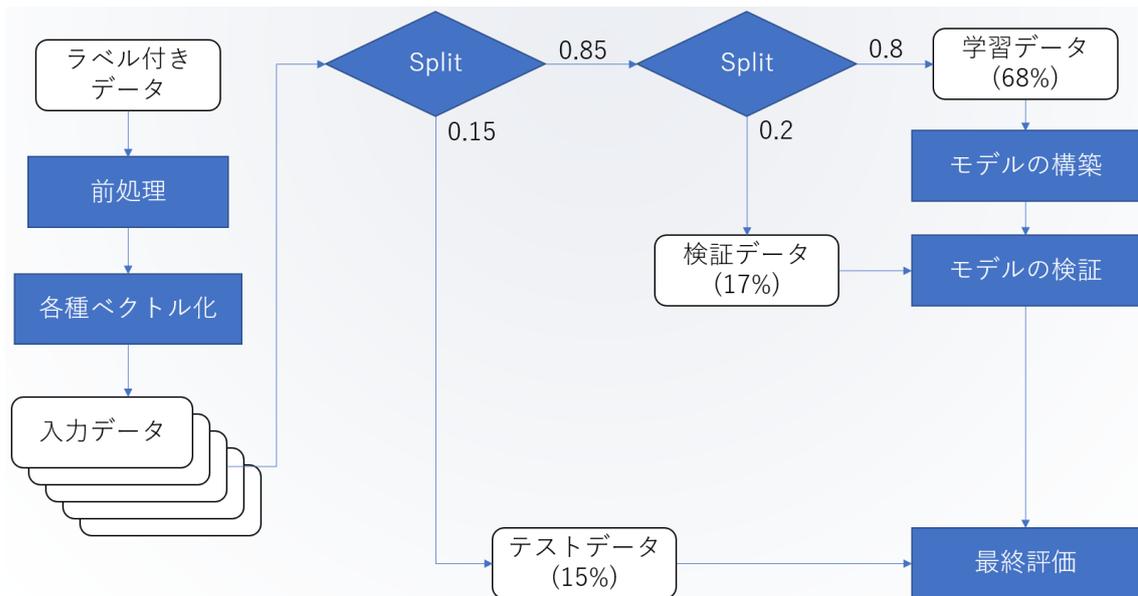


図 A2-1-4-1

A2.1.4.1. コサイン類似度を用いて判定するモデル

作成したプログラムの評価結果は表 A2-1-4-1-1 の通りであった。

表 A2-1-4-1-1

目的変数	F1	Recall	Accuracy	Precision
LABEL1	0.835	0.819	0.839	0.853

A2.1.4.2. アンサンブルモデル

作成したアンサンブルモデルの評価結果は表 A2-1-4-2-1 の通りであった。

表 A2-1-4-2-1

目的変数	F1	Recall	Accuracy	Precision
LABEL1	0.877	0.987	0.862	0.789
LABEL2	0.804	0.940	0.900	0.703

また、アンサンブルに使用した各個別モデルの評価結果は表 A2-1-4-2-2 の通りであった

25。

表 A2-1-4-2-2

目的変数	Input -手法	F1	Recall	Accuracy	Precision
LABEL1	BoW - LR	0.909	0.893	0.911	0.925
	TF-IDF - SVC	0.928	0.939	0.927	0.917
	Emb - LSVC	0.896	0.888	0.898	0.905
	TF-IDF × Emb - NN	0.868	0.891	0.865	0.846
	BERT	0.931	0.920	0.932	0.943
LABEL2	BoW - SVC	0.857	0.831	0.940	0.885
	TFIDF - LR	0.813	0.735	0.927	0.910
	Emb - RR	0.745	0.687	0.898	0.814
	TF-IDF × Emb - NN	0.657	0.554	0.874	0.807
	BERT	0.727	0.675	0.890	0.789

今回の検証において作成した各個別モデルの最良 F1 スコアは表 A2-1-4-2-3 および表 A2-1-4-2-4 の通りであった。

表 A2-1-4-2-3 (LABEL1 を使用したスコア)

	BoW	TFIDF	Emb	TF-IDF × Emb	BERT
LR	0.909	0.926	0.895	0.826	
RR	0.900	0.923	0.881	0.780	
SVC	0.901	0.928	0.895	0.845	
LSVC	0.907	0.927	0.896	0.820	
NB	0.877	0.877	0.753	0.735	
KN	0.812	0.850	0.835	0.828	
RF	0.842	0.853	0.841	0.854	
AB	0.885	0.889	0.883	0.855	
LGBM	0.875	0.871	0.873	0.866	
NN	0.907	0.897	0.897	0.868	
BERT					0.931

25 表中の略語は以下の通り：

BoW: Bag of Words、LR: ロジスティック回帰、RR: リッジ回帰、LSVC: 線形 SVC、NB: ナイブベイズ、KN: k 近傍法、RF: ランダムフォレスト、AB: Ada Boost、LGBM: Light GBM、NN: ニューラルネットワーク、Emb: 分散表現

表 A2-1-4-2-4 (LABEL2 を使用したスコア)

	BoW	TFIDF	Emb	TFIDF × Emb	BERT
LR	0.828	0.813	0.725	0.578	
RR	0.840	0.765	0.745	0.291	
SVC	0.857	0.800	0.747	0.544	
LSVC	0.828	0.795	0.730	0.566	
NB	0.694	0.694	0.525	0.391	
KN	0.522	0.650	0.550	0.437	
RF	0.092	0.114	0.556	0.452	
AB	0.753	0.708	0.703	0.515	
LGBM	0.658	0.644	0.676	0.576	
NN	0.793	0.758	0.725	0.657	
BERT					0.727

A2.1.5. 考察

A2.1.5.1. 評価結果について

LABEL1 を目的変数とした分類においては、コサイン類似度を用いて判定するモデルでは F1 スコアが 0.835、Recall が 0.819 であったのに対し、アンサンブルモデルでは F1 スコアが 0.877、Recall が 0.987 と、機械学習を用いたフリーテキストの判定において機械学習が一定の有用性をもつ可能性が示唆された。一方で LABEL2 を目的変数とした分類では、アンサンブルモデルにおいて F1 スコアが 0.804、Recall が 0.940 と、LABEL1 を使用した時ほどの分類精度は出ていない。この主な原因としては、LABEL2 の学習データは LABEL1 の学習データの半量となる約 2500 件しかなく、その中で Positive ラベルが 635 件と学習データの 1/4 しかないため、適切な分類を行うための十分な学習ができなかった可能性やデータの偏りが発生している可能性が考えられる。

A2.1.5.2. 入力値とするベクトルについて

個別のモデルに目を移すと、Bag of Words や TF-IDF を使用するモデルが分散表現を使用するモデルよりもスコアが高い傾向にあった。Bag of Words や TF-IDF においては単語に注目した判定を行うことができた一方で、分散表現を用いた分類が意図したほど効果を上げていないことがうかがえる。原因としては、①今回対象としている文章は短文が多く分散表現による効果を得られにくかったこと、②参照した事前学習モデルから適切な分散表現が得られなかったこと、③分散表現を使用した分類を行うにはデータ量が不十分であっ

た、などの理由が考えられる。分散表現の取得は東北大学のモデルを使用させていただいたが、事前学習モデルから取得できない単語のベクトルはゼロとなる。そのため、一般的な用語が多く含まれる Wikipedia で事前学習されたモデルよりも、医学ドメインに特化した事前学習モデルを使用、もしくは新たに作成することで判定性能の向上が見込める可能性がある。これは BERT も同様であり、医学用語を含む文章で事前学習された MEDPERTjp [68]や UTH-BERT [69]を使用することでモデルの判定精度が向上する可能性がある。

なお、今回用意した訓練データにおける Bag of Words や TF-IDF の次元は約 5000 次元であり、次元削減を行わずに処理を行っている。これらのベクトル化手法ではデータ量とともに次元が増え、学習効率や判定精度にも影響するため、データを増やす場合は必要に応じて次元削減などの対応も検討する必要がある。

A2.1.5.3. 前処理について

テキストの前処理は AI による判定に大きく影響するため、慎重な検討が必要となる。不要な語彙は判定モデルの作成においてノイズとなるため、表記揺れの正規化や辞書を用いた同義語の統一、判定に寄与しない語彙の削除などを行うことで判定精度の向上が見込める。そのためには、適切な形態素解析が行えることが必要となる。例えば、臨床検査値異常として報告される「高値」や「低値」という用語は、今回のプロトタイプで用いた設定（形態素解析器：MeCab、システム辞書：NEologd、ユーザー辞書：万病辞書）では以下のように形態素解析される。

高値	高値	名詞,一般,*,*,*,高値,タカネ,タカネ
低値	低 値	接頭詞,名詞接続,*,*,*,低,テイ,テイ 名詞,一般,*,*,*,値,ネ,ネ

読み方は異なるが「高値」がそのまま 1 単語として維持されているのに対し、「低値」は 2 単語に分割されており、「低」が接頭詞の扱いとなっている。そのため、モデルに読み込ませる対象として残した 4 つの品詞（名詞、形容詞、助動詞、動詞）に含まれず、今回のプロトタイプでは「低」という情報が活用されていない。このように意図しない分かち書きを防ぐためにも、意図した形態素に分割されるよう形態素解析器が設定されていることの確認が欠かせない。プロトタイプでは品詞に着目して語彙の絞り込みを行ったが、品詞による絞り込みでは上記のような意図しない除外が発生してしまう可能性があるため、不要とできる単語（ストップワード）を定義して語彙の絞り込みを行うことで、より意義のある単語を用いた判定モデルの作成が見込める。

また、今回のプロトタイプでは同義語の統一は行っていない。各社で使用している辞書を用いて医学事象や薬剤名の統一を行うことで、より判定精度の高いモデルを作成できる可能性がある。

A2.2. プロジェクトへの実装に向けて

今回のプロトタイピングでは、機械学習をデータマネジメントの実務に使えるか試してみよう、という思いから進められており、プログラムの処理速度や運用におけるメンテナンスなどの対応は一切考慮されていない。そのため、実際のプロジェクトに実装するためにはAIの専門家の助力を仰ぎながら、実際の運用に耐えうる処理速度や判定結果が得られるようにプログラムを設計し、継続して運用していくための手順も検討していく必要がある。しかしながら、AIの専門家に我々が抱えている課題を適切に伝え、また、専門家からの提案を正しく理解して運用するためには、AIがどのように動いているのか、どのような点が弱点になるのかを理解しておくことは欠かせないであろう。実務に耐えうるAIやシステムを構築することはできずとも、プロトタイピングを通じてAIの仕組みに対する理解を深めることができるため、プログラミングに興味のある方は挑戦してみるのはいかがでしょうか。

参考文献

- [1] 統合イノベーション戦略推進会議, “A I 戦略 2022,” 22 4 2022. Available: https://www8.cao.go.jp/cstp/ai/aistrategy2022_honbun.pdf.
- [2] Demetris N Zambas (Pfizer), “Smart Data Query - Challenge to Apply Machine Learning,” 26th DIA Japan Annual Workshop for Clinical Data Management, 2023.
- [3] N. P. K. Mirai Kikawa and Yuichi Nakajima, “How to let Machine Learn Clinical Data Review as it can Support Reshaping the Future of Clinical Data Cleaning Process,” PharmaSUG, 2020.
- [4] Marilina Castellano (Uppsala Monitoring Centre), “A Practical Example of AI Supported Medical Coding,” 26th DIA Japan Annual Workshop for Clinical Data Management, 2023.
- [5] 内閣府等, “「AI 人材獲得・人材育成及び研究環境整備に関するアンケート」集計結果,” 2020. Available: https://www8.cao.go.jp/cstp/ai/ai_q_2.pdf.
- [6] 内閣府, “人間中心の AI 社会原則,” 29 Mar 2019. Available: <https://www.cas.go.jp/jp/seisaku/jinkouchinou/pdf/aigensoku.pdf>.
- [7] EMA, “Artificial intelligence in medicine regulation News 16/08/2021,” 16 Aug 2021. Available: <https://www.ema.europa.eu/en/news/artificial-intelligence-medicine-regulation>.
- [8] U. M. Centre, “WHODrug Koda,” Available: <https://who-umc.org/whodrug/whodrug-koda/>.
- [9] “WHODrug Koda - Your new coding powerhouse,” Uppsala Monitoring Centre, 2022. Available: <https://www.youtube.com/watch?v=6GaxSoT6mX0&t=52s>.
- [10] “WHODrug Koda - a complete overview, including the latest updates and vendor experiences,” Uppsala Monitoring Centre, Available: https://www.youtube.com/watch?v=c8_QWjssFyo.
- [11] U. M. Centre, “WHODrug Webinar October 2022 - Koda Webinar,” Available: https://who-umc.org/media/py2a2kt4/whodrug-koda-webinar_umc.pdf.
- [12] Rao, Prasanna (Pfizer), “AI and Machine Learning: What it is and its Impact on Data Management,” SCDM 2022 Annual Conference, San Antonio, Texas, 2022.
- [13] Takeshi Kawakami (Pfizer), “Smart Data Query - Challenge to apply machine learning technology,” 26th DIA Japan Annual Workshop for Clinical Data Management, 2023.
- [14] 文部科学省, “数理・データサイエンス・AI 教育プログラム認定制度（リテラシーレベル）,” Available: https://www.mext.go.jp/a_menu/koutou/suuri_datascience_ai/00002.htm.

- [15] 数理・データサイエンス・AI教育プログラム認定制度検討会議, “「数理・データサイエンス・AI教育プログラム認定制度（リテラシーレベル）」の創設について,” 3月2020年. Available: <https://www8.cao.go.jp/cstp/ai/suuri/ninteisousetu.pdf>.
- [16] 日本製薬工業協会 データサイエンス部会, “機械学習始めました,” 日本製薬工業協会, 2020.
- [17] K. Huysentruyt, O. Kjoersvik, P. Dobracki, E. Savage, E. Mishalov, M. Cherry, E. Leonard, R. Taylor, B. Patel, D. Abatemarco, “Validating Intelligent Automation Systems in Pharmacovigilance: Insights from Good Manufacturing Practices,” 2021.
- [18] 経済産業省, “我が国のAIガバナンスの在り方 ver. 1.1,” 9 Jul 2021. Available: https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20210709_1.pdf.
- [19] OECD.AI, “OECD AI Principles overview,” May 2019. Available: <https://oecd.ai/en/ai-principles>.
- [20] N. A. H. H. A. C. N. M. S. Jessica Fjeld, “Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI” Berkman Klein Center for Internet & Society,” Berkman Klein Center Research Publication, 2020.
- [21] 欧州委員会, “WHITE PAPER_On Artificial Intelligence_A European approach to excellence and trust,” 19 Feb 2020. Available: https://commission.europa.eu/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en.
- [22] 欧州委員会, “Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS,” 21 Apr 2021. Available: https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0001.02/DOC_1&format=PDF.
- [23] 米国食品医薬品局（FDA）、カナダ保健省、英国医薬品医療製品規制庁（MHRA）, “Good Machine Learning Practice for Medical Device Development: Guiding Principles,” Oct 2021. Available: <https://www.fda.gov/medical-devices/software-medical-device-samd/good-machine-learning-practice-medical-device-development-guiding-principles>.
- [24] FDA, “Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices,” 5 Oct 2022. Available: <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices>.
- [25] 総務省, “AI利活用ガイドライン,” 9 Aug 2019. Available: https://www.soumu.go.jp/main_content/000637097.pdf.

- [26] 経済産業省, “AI・データ利用に関する契約ガイドライン 1.1 版 AI 編,” Jun 2018. Available:
https://www.meti.go.jp/policy/mono_info_service/connected_industries/sharing_and_utilization/20180615001-3.pdf.
- [27] 経済産業省, “AI・データ利用に関する契約ガイドライン 1.1 版データ編,” Jun 2018. Available:
https://www.meti.go.jp/policy/mono_info_service/connected_industries/sharing_and_utilization/20180615001-2.pdf.
- [28] 日本製薬工業協会 医薬品評価委員会 臨床評価部会 特別プロジェクト3, “医薬品開発及びデータ二次利用における個人情報保護に関する留意点,” Apr 2022. Available:
https://www.jpma.or.jp/information/evaluation/results/allotment/rfcmr000000028kz-att/privacy_points_remember_202204.pdf.
- [29] 総務省, “国際的な議論のための AI 開発ガイドライン案,” 28 Jul 2017. Available:
https://www.soumu.go.jp/main_content/000499625.pdf.
- [30] Information Commissioner’s Office (ICO) and The Alan Turing Institute, “Explaining decisions made with AI,” 20 May 2020. Available: <https://ico.org.uk/media/for-organisations/guide-to-data-protection/key-dp-themes/explaining-decisions-made-with-artificial-intelligence-1-0.pdf>.
- [31] 経済産業省, “AI 導入ガイドブック 構想ステージ,” 2022. Available:
https://www.meti.go.jp/policy/it_policy/jinzai/1_AI_guidebook_mihiraki_FIX.pdf.
- [32] Forbes Business Council, “How Small-Business Owners Can Adopt AI Solutions With Limited Resources,” 2022. Available:
<https://www.forbes.com/sites/forbesbusinesscouncil/2022/12/21/how-small-business-owners-can-adopt-ai-solutions-with-limited-resources/?sh=4daf65ce7df0>.
- [33] G. Kesari, “You’ve Invested In AI, But Are You Getting ROI From It?,” 2021. Available: <https://www.forbes.com/sites/ganeskesari/2021/03/29/youve-invested-in-ai-but-are-you-getting-roi-from-it/?sh=3b0797e36063>.
- [34] 総務省, “ICT スキル総合習得教材,” Available:
https://warp.da.ndl.go.jp/collections/content/info:ndljp/pid/12475307/www.soumu.go.jp/ict_skill/pdf/ict_skill_3_5.pdf.
- [35] 総務省統計局, “「高等学校における「情報 II」のためのデータサイエンス・データ解析入門」,” Available: <https://www.stat.go.jp/teacher/comp-learn-04.html>.
- [36] 東京大学, “UTokyo OpenCourseWare,” Available: <https://ocwx.ocw.u-tokyo.ac.jp/>

tokyo.ac.jp/course_11414/.

- [37] kaggle, Available: <https://www.kaggle.com/>.
- [38] Papers with code, Available: <https://paperswithcode.com/>.
- [39] Paper with code, “Datasets,” Available: <https://paperswithcode.com/datasets>.
- [40] 東京大学, “Python プログラミング入門,” Available: <https://utokyo-ipp.github.io/>.
- [41] 三谷 純 (筑波大学) , Available:
https://mitani.cs.tsukuba.ac.jp/book_support/python/python_slides.pptx.
- [42] Google LLC, “Google Colaboratory,” Available: <https://colab.research.google.com/>.
- [43] Coursera, “Supervised Machine Learning: Regression and Classification,” Available:
<https://www.coursera.org/learn/machine-learning>.
- [44] J. L. R. L. Z. a. W. Z. Sheela Kolluri, “Machine Learning and Artificial Intelligence in Pharmaceutical Research and Development: a Review,” 2022.
- [45] P. M. M. N. C. f. H. S. Christine A. Lucas , M. R. C. M. J. N. M. P. B. M. M. R. T. M. D. M. P. M. C. C. M. R. I. Emily Hadley, “Machine Learning for Medical Coding in Healthcare Surveys,” 2021.
- [46] S.-W. C. , T. Y. Takuma Oda, “Semi-automated Conversion of Clinical Trial Legacy Data into CDISC SDTM Standards Format Using Supervised Machine Learning,” 2021.
- [47] T. N. T. A. R. E. Y. L. D. F. F. J. B. C. H. F. K. S. S. R. E. E. Hope Weissler, “The role of machine learning in clinical research: transforming the future of evidence generation,” 2021.
- [48] K. A. T. & Ł. Kidziński, “Artificial intelligence can improve patients’ experience in decentralized clinical trials, Nature Medicine volume 28, pages2462-2463 (2022),” 2022.
- [49] 日本製薬工業協会 データサイエンス部会, “AI ってなに?,” 日本製薬工業協会, 2019.
- [50] “Some studies in machine learning using the game of checkers,” IBM, Available:
<https://ieeexplore.ieee.org/document/5389202/metrics>.
- [51] “javatpoint Decision Tree Classification Algorithm,” Available:
<https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm>.
- [52] 周. 志華, 著: 機械学習, 近代科学社, 2022.
- [53] “What is random forest?,” BMI, Available: <https://www.ibm.com/topics/random-forest>.
- [54] “javatpoint Logistic Regression in Machine Learning,” Available:
<https://www.javatpoint.com/logistic-regression-in-machine-learning>.
- [55] “javatpoint K-Nearest Neighbor(KNN) Algorithm for Machine Learning,” Available:

- <https://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machine-learning>.
- [56] “javatpoint Naïve Bayes Classifier Algorithm,” Available:
<https://www.javatpoint.com/machine-learning-naive-bayes-classifier>.
- [57] “javatpoint Artificial Neural Network Tutorial,” Available:
<https://www.javatpoint.com/artificial-neural-network>.
- [58] “javatpoint Support Vector Machine Algorithm,” Available:
<https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm>.
- [59] 山本和英, 実践・自然言語処理シリーズ テキスト処理の要素技術, 近代科学社, 2021.
- [60] 黒橋禎夫, 改訂版 自然言語処理, NHK 出版, 2019.
- [61] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” 2019.
- [62] 厚生労働省, “厚生科学審議会 (予防接種・ワクチン分科会 副反応検討部会),” Available: https://www.mhlw.go.jp/stf/shingi/shingi-kousei_284075.html.
- [63] LINE Corp., “mecab-ipadic-NEologd : Neologism dictionary for MeCab,” 14 Sep 2020. Available: <https://github.com/neologd/mecab-ipadic-neologd>.
- [64] 奈良先端科学技術大学院大学 ソーシャル・コンピューティング研究室, “万病辞書 Manbyo-Dictionary,” Available: <https://sociocom.naist.jp/manbyou-dic/>.
- [65] 東北大学 自然言語処理研究グループ, “Wikipedia Entity Vectors,” 11 May 2021. Available: <https://github.com/singletonue/WikiEntVec>.
- [66] 東北大学 自然言語処理研究グループ, “Pretrained Japanese BERT models,” 26 Aug 2021. Available: <https://github.com/cl-tohoku/bert-japanese>.
- [67] M. T. C. Ribeiro, “lime,” 30 Jul 2021. Available: <https://github.com/marcotcr/lime>.
- [68] 大阪大学医学部附属病院 医療情報部, “Trials of pre-trained BERT models for the medical domain in Japanese,” 22 Nov 2020. Available: <https://github.com/ou-medinfo/medbertjp>.
- [69] 東京大学大学院 医学系研究科 医療 AI 開発学講座, “UTH-BERT: a BERT pre-trained with Japanese clinical text,” Available: <https://ai-health.m.u-tokyo.ac.jp/home/research/uth-bert>.

タスクフォースメンバー

EA ファーマ株式会社	橘田 由香里
グラクソ・スミスクライン株式会社	久田 大輔 (リーダー)
ノバルティス ファーマ株式会社	井上 祐介
ファイザーR&D 合同会社	三沢 秀敏 (リーダー)
ファイザーR&D 合同会社	小宮山 靖 (担当副部会長)
ファイザーR&D 合同会社	田中 拓海
ヤンセンファーマ株式会社	二瓶 浩美
大塚製薬株式会社	山川 ひろみ
株式会社ヤクルト本社 (～2023年2月)	柴田 奈津実
アステラス製薬株式会社 (2023年3月～)	

レビューワー

エーザイ株式会社	酒井 弘憲
グラクソ・スミスクライン株式会社	内海 啓介
サノフィ株式会社	加藤 智子
ファイザーR&D 合同会社	土綿 慎一
住友ファーマ株式会社	土屋 悟
中外製薬株式会社	坂上 拓
中外製薬株式会社	山本 英晴
中外製薬株式会社	淡路 直人
日本ベーリンガーインゲルハイム株式会社	長谷川 秀美