データマネジメントにおけるAI（Artificial Intelligence）の活用を考えてみる

東京大学医学部附属病院
臨床研究推進センター
内海　啓介

はじめに

本演題発表は，演者が前職で活動していた日本製薬協 医薬品評価委員会 データサイエンス部会での活動の成果[1),2)]を基にしたものであり，私が所属する東京大学医学部附属病院 臨床研究推進センターを代表するものではありません.

1) 日本製薬工業協会 データサイエンス部会, “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 ～,” 5 月 2023. [オンライン]. Available: 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 ～ | 医薬品評価委員会の成果物 一覧 | 日本製薬工業協会 (jpma.or.jp)
2) 日本製薬工業協会 データサイエンス部会, “データマネジメント業務におけるAI活用検討のためのモデル試作,” 6月 2024. [オンライン]. Available: データマネジメント業務におけるAI活用検討のためのモデル試作 | 医薬品評価委員会の成果物 一覧 | 日本製薬工業協会 (jpma.or.jp)

データマネジメントにおけるAIの活用

AIは膨大なデータを効率的に処理し，精度の高い分析を行うためにますます重要な技術となっており，AIの活用は今後ますます広がっていくと考えられます．

European Medicines Agency といった規制当局もそのReflection paperの中で，「AI と機械学習ツールは，正しく使用されれば，医薬品ライフサイクル内でのデータの取得，変換，分析および解釈を効果的にサポートすることができる」と AI の当該分野での有用性を認めています[3)]．

臨床開発から製造販売後調査に至るまで，データマネジメントのあらゆる段階でAIの役割が増大することが予想されます．

データマネジメント分野においてAIがどのように活用されるかを、具体的な事例を交えて考えてみます．

3) European Medicines Agency, “Reflection paper on the use of AI in the medicinal product lifecycle, ” July 2023. [オンライン]. Available: https://www.ema.europa.eu/en/documents/scientific-guideline/draft-reflection-paper-useartificial-intelligence-ai-medicinal-product-lifecycle_en.pdf.

ユースケース

今回はこの3つの点について触れます

データマネジメントにおけるAI利用の主なユースケースとして，以下の例が挙げられます．

- **コーディング**：CRF（Case Report Form，症例報告書）に報告した用語（有害事象，併用薬等）を標準的な用語に結び付ける．DMのリソースと時間の軽減が可能となる．
- **データレビューとクエリ文作成**：CRFデータの不整合を検出し，自動的にクエリ文を生成するプロセスを導入する．DMのリソースの軽減に加えて，データレビューのサイクルタイムが短縮され，施設での対応やモニタリング業務への負荷の軽減，データベースの早期固定に寄与できる．
- **併用薬剤**の使用目的の妥当性の判断．DMのリソースと時間の軽減が可能となる．
- **臨床試験**：eCoAへの入力タスクの通知最適化や画像データの適切性フィードバック，少数のセンサーでの詳細な測定などへの利用．

ユースケース

- **コーディング**: CRF（Case Report Form，症例報告書）に報告した用語（有害事象，併用薬等）を標準的な用語に結び付ける.
- **データレビューとクエリ文作成**: CRFデータの不整合を検出し，自動的にクエリ文を生成するプロセスを導入する.
- **併用薬剤**の使用目的の妥当性の判断.

コーディング

コーディング

コーディングとは，CRF で報告され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.

腹痛

入力・転記

EDC/
紙CRF

クエリー

データベース

10000081 腹痛

自動

不一致となった
報告

辞書

手動

データマネージャ，コーダー

コーディング用の**辞書**

- **有害事象** → MedDRA
- **併用薬** → WHO-DD，各企業の辞書

辞書でコーディングが完了しない場合

- データマネージャ，あるいはコーダーが報告用語を確認し、クエリーを出して解決するか、マニュアル等コーディングを実施
- 辞書でコーディングが完了できるようにシステムやシノニムリストをupdateするケースもあり，各組織でのノウハウで対応

UMC (Uppsala Monitoring Center)*の事例[4), 5), 6), 7)]

報告された薬剤名：

naphazoline (naphcon) 0.1% ophthalmic solution

投与経路：
conjunctival

適応：
glaucoma

WHODrug Koda

コーディングされた薬剤名：
NAPHCON

有効成分：
NAPHAZOLINE HYDROCHLORIDE

ATCコード**：
S01GA, SYMPATHOMIMETICS USED AS DECONGESTANTS

* **UMC (**Uppsala Monitoring Centre)：国際的な医薬品安全性における科学的研究とサービスを提供する独立非営利財団法人
****ATC**（Anatomical Therapeutic Chemical Classification System）：解剖治療化学分類法．医薬品統計法共同研究センター（Collaborating Centre for Drug Statistics Methodology）によって管理されている医薬品分類方法の1つ

WHO個別症例安全性報告

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がサポートしている
WHOデータベースを学習
データに利用（数百万件）

10,000個

WHODrug Koda

コーディング（薬剤名，ATCコード）

153 時間

558 時間
（人のみが実施）

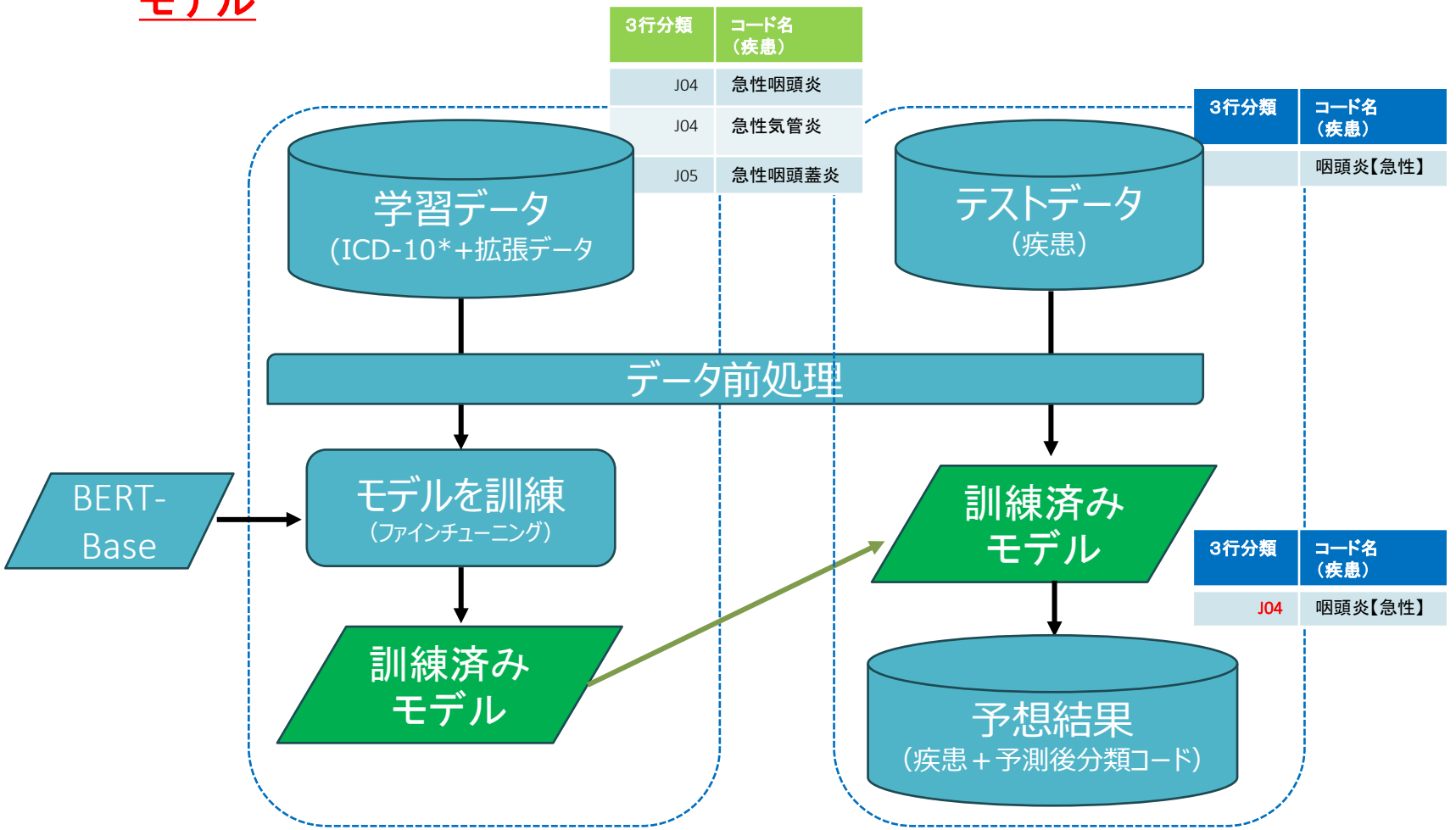
4) U. M. Centre, “WHODrug Koda,” Available: https://who-umc.org/whodrug/whodrugkoda/.
5) “WHODrug Koda - Your new coding powerhouse,” Uppsala Monitoring Centre, 2022. Available: https://www.youtube.com/watch?v=6GaxSoT6mX0&t=52s.
6) “WHODrug Koda – a complete overview, including the latest updates and vendor experiences,” Uppsala Monitoring Centre, Available: https://www.youtube.com/watch?v=c8_QWjssFyo.
7) U. M. Centre, “WHODrug Webinar October 2022 -Koda Webinar,” Available: https://who-umc.org/media/py2a2kt4/whodrug-koda-webinar_umc.pdf.

製薬協のTFメンバーで試してみた　– コーディング　-

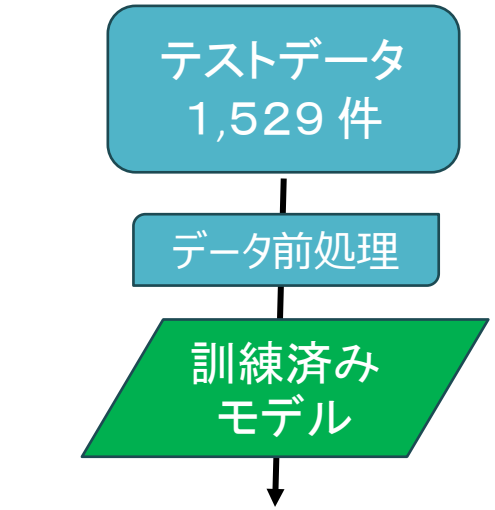
- コーディングを複数のカテゴリから1つを選択する分類タスクと捉え，テキスト分類に用いることができるBERTモデル*を利用．学習データで本モデルを訓練させて，テストデータで疾患名のコーディング予測結果をTFメンバーで評価．

＊ICD（International Statistical Classification of Diseases and Related Health Problems）：　疾病および関連保健問題の国際統計分類．異なる国や地域から、異なる時点で集計された死亡や疾病のデータの体系的な記録，分析，解釈および比較を行うため，世界保健機関憲章に基づき，世界保健機関（WHO）が作成した分類

モデル



結果



正解率		
Top 1	Top 3	Top 5
82.4 %	90.7 %	92.4 %

Top 1とは、モデルが予測した分類コードの予測結果の中で最も高い確率で示された分類コード．Top 3とTop 5は、それぞれ予測結果の中で高い確率が示された上位3つ，上位5つの分類コードを示す．

＊BERTモデル：　Bidirectional Encoder Representations from Transformersモデルの略称で，Googleが発表した自然言語処理のモデル

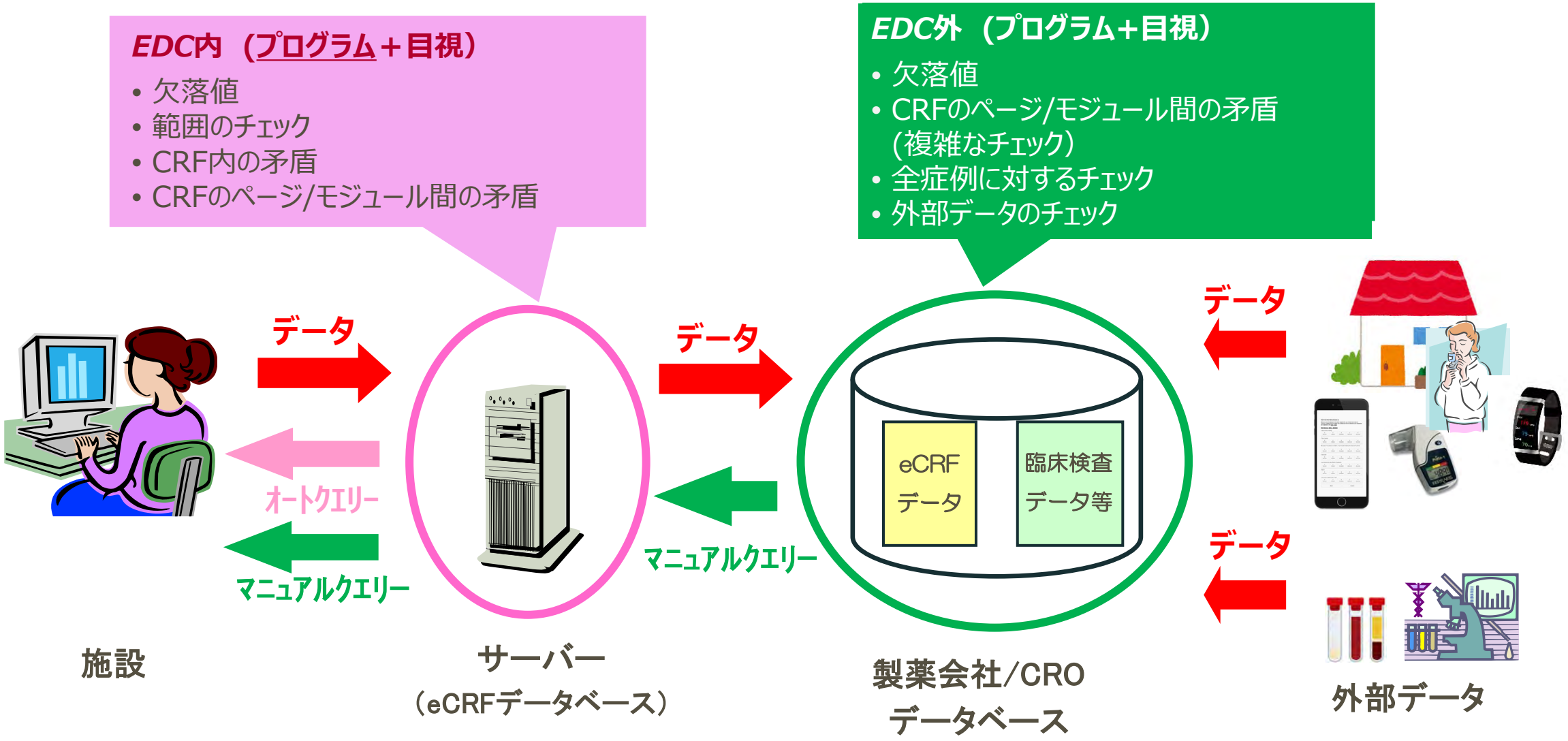
コーディングのまとめ

- 学習用データとして標準コードデータしかなかったものの，辞書分類コード を目的とした分類において，機械学習モデルを利用した分類が一定の効果を発揮することが確認できた.
- MedDRAはライセンスの都合上用いることはできなかったためICD-10を用いたが，MedDRAのLLTコードを予測するモデルにも，本プロトタイピングの構築手法を適用することは可能であると考えられた． その際には，各社・各病院で蓄積された報告語のコーディングデータを学習データとして用いることが可能である.
- すでに実用化されているAIシステムがあること，企業で用いているコーディングシステムからAIの適応は予想はできていたものの，今回の検討の結果から，リスクと精度の観点から必要と思われるレベルの人の介在を前提に，現在の人による作業の一部または大部分をAIに置き換えることで一定の成果が上げられると考えられた.

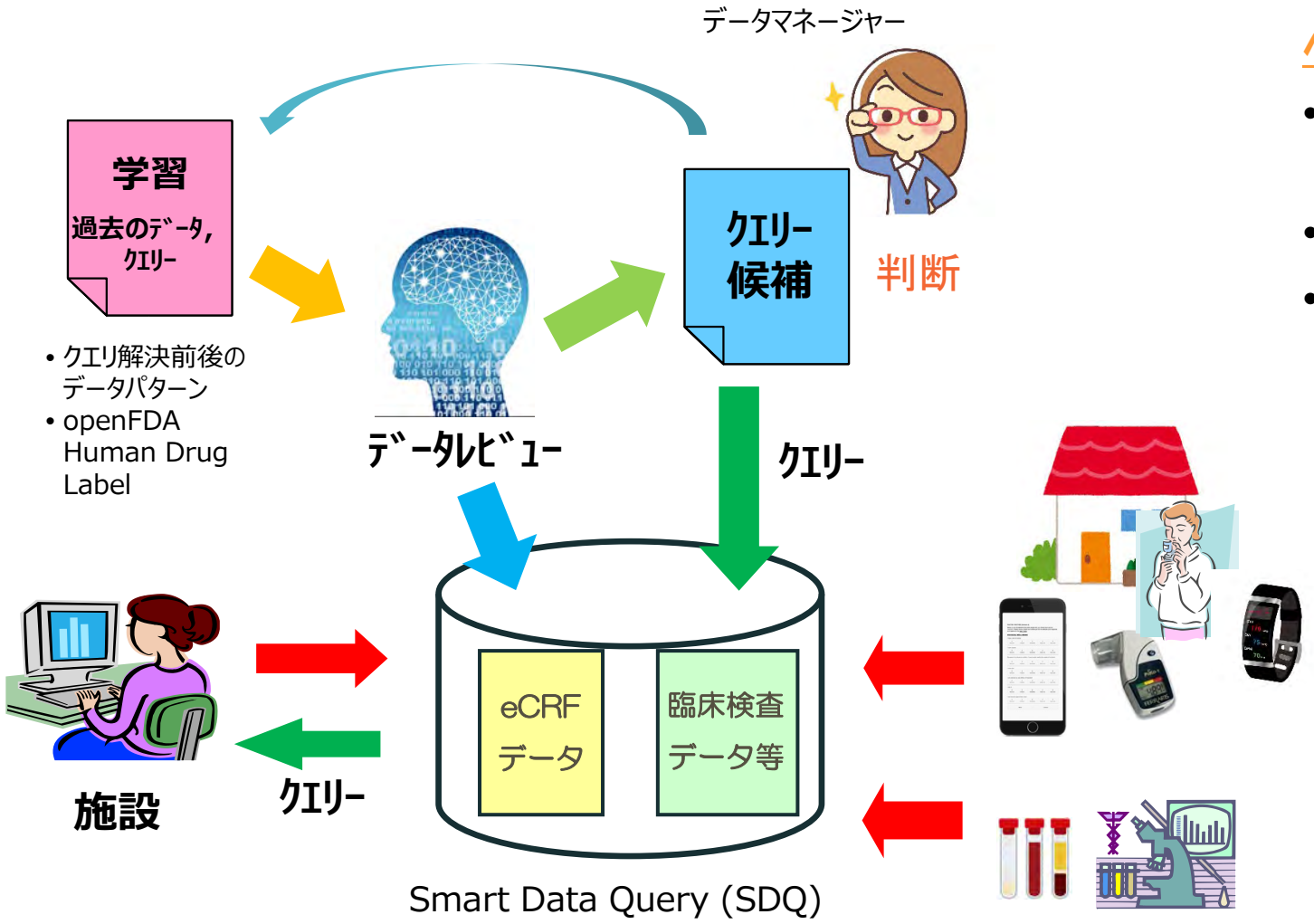
ユースケース

- **コーディング**: CRF（Case Report Form，症例報告書）に報告した用語（有害事象，併用薬等）を標準的な用語に結び付ける.
- **データレビューとクエリ文作成**: CRFデータの不整合を検出し，自動的にクエリ文を生成するプロセスを導入する.
- **併用薬剤**の使用目的の妥当性の判断.

データレビューとクエリー文の作成　（EDCの例）



製薬会社のAIを用いた事例[8), 9)]



ベネフィット

- 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
- データレビューおよびクエリー作成の労力削減
- リアルタイムに近いクエリ発行

新型コロナワクチンの治験では，従来約27日かかったクエリ解決の期間が，1.7日に短縮

8) Rao, Prasanna (Pfizer), “AI and Machine Learning: What it is and its Impact on Data Management,” SCDM 2022 Annual Conference, San Antonio, Texas, 2022.
9) Takeshi Kawakami (Pfizer), “Smart Data Query - Challenge to apply machine learning technology,” 26th DIA Japan Annual Workshop for Clinical Data Management, 2023.

製薬会社のAIを用いた事例[8), 9)]

例）有害事象と併用薬の整合性確認

有害事象に対して併用薬が使用された場合

有害事象により**併用薬**が使用された
→ はい
その有害事象の発現に対応する投与開始日/終了日を持つ併用薬はあるか？
→ はい
その有害事象に対する治療と整合する併用薬はあるか？
→ はい
併用薬の使用目的として特定されている有害事象は当該有害事象と一致しているか？
→ いいえ
クエリーがEDCに掲載され、施設に送られる

例示

AE ID：1

AE Term: FEVER

Was concomitant medication given: YES

Onset date： 2023/02/01
Resolved date: 2023/02/02

有害事象

＋

CM ID：1

CM Term: Loxoprofen Sodium Hydrate

Start date： 2023/02/01
Stop date: 2023/02/02

AE ID: (Blank)

併用薬

→

“Was a concomitant Medication given?” is YES: however, there is a potentially corresponding medication reported [#1, Loxoprofen Sodium Hydrate, 2023/02/01-2023/02/02] that is not reported in the AE identifier field. Please clarify or update if appropriate.

クエリー文

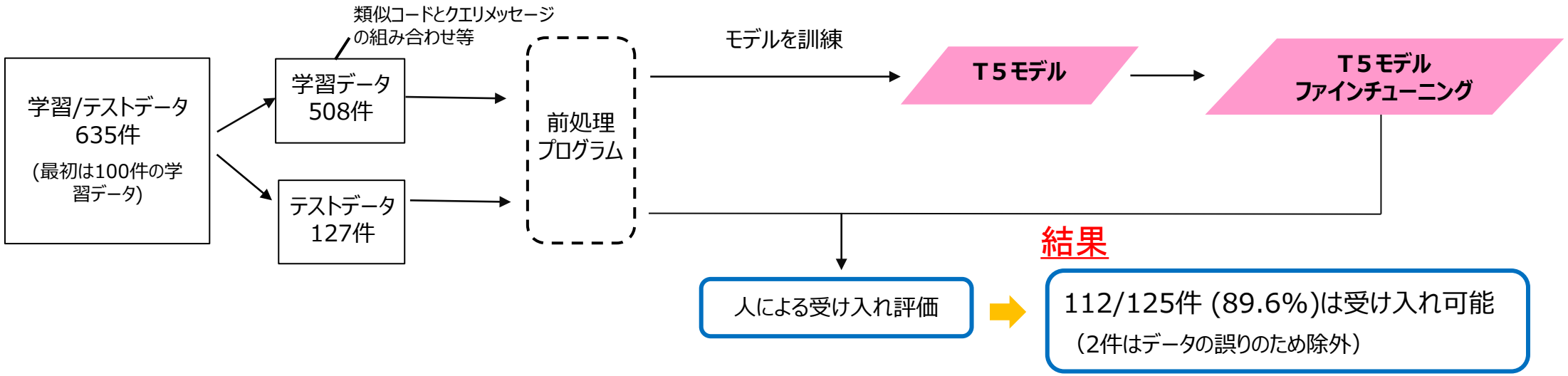
施設

製薬協のTFメンバで試してみた　– クエリ-文作成　-

- EDCに入力されたデータが論理的に正しいかどうかをチェックする論理チェック仕様を元に，クエリメッセージ文案を自動で作成することが可能であるかどうかを検討した．期待は，論理チェック毎にクエリー文をマニュアルで準備をする作業を軽減すること．
- ファインチューニングによるアプローチおよび大規模言語モデルを用いたアプローチの2通りの方法を用いて検討した．

ファインチューニングアプローチ

仕様からメッセージの変換を言語翻訳と捉え，翻訳等に用いることができるT5モデル*を用いた．

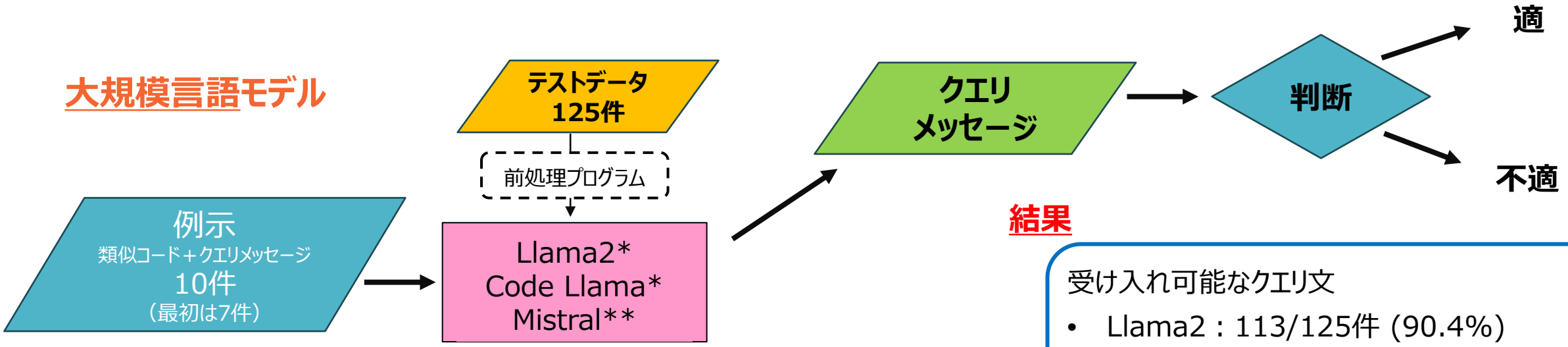


- 比較的高い精度にはなったものの，改善のためには，さらに良い学習データが必要．各組織（企業）内の実績データを学習データとして使うことは有効．
- 受け入れ不可を補うような学習データばかりを増やすのではなく，バランスよく学習データを増やすことが必要．

疑似コード	期待したメッセージ	T5（ファインチューニング）の結果	受け入れ不可理由
[Start Date of Radiation Therapy] >= [Date of First Study Drug Taken	[Start Date of Radiation Therapy] is equal to or after the [Date of First Study Drug Taken] on the randomization form. please review	[Start Date of Radiation Therapy] is after the [Date of First Study Drug Taken] on the study drug dosage form. please review.	イコールの要素が欠落している

＊T5モデル：Text-to-Text Transfer Transformerモデルの略称で、質問応答、翻訳、要約など、テキストからテキストへの変換をおこなうことができる、Googleが発表した自然言語処理のモデル

製薬協のTFメンバで試してみた　– クエリ-文作成　-



例（類似コードが複雑だとクエリー文も長文）

疑似コード	期待する出力メッセージ	Llama2
[Is the adverse event serious?] = "YES" AND [Congenital Anomaly or Birth Defect] = "NO" AND [Persist or Signif Disability/Incapacity] = "NO" AND [Did this serious event result in death?] = "NO" AND [Requires or Prolongs Hospitalization] = "NO" AND [Is Life Threatening] = "NO" AND [Other Medically Important Serious Event] = "NO"	[Is the adverse event serious?] is marked as "YES", but responses to all of the seriousness criteria are NO. At least one seriousness criterion is expected to be "YES" for serious events. Please review and update as appropriate	The [Is the adverse event serious?] is "YES", yet the [Congenital Anomaly or Birth Defect] is "NO" and the [Persist or Significant Disability/Incapacity] is "NO" and the [Did this serious event result in death?] is "NO" and the [Requires or Prolongs Hospitalization] is "NO" and the [Is Life Threatening] is "NO" and the [Other Medically Important Serious Event] is "NO". Please review

結果

受け入れ可能なクエリ文
- Llama2：113/125件 (90.4%)
- Code Llama：111/125件 (88.8%)
- Mistral：106/125 (84.8%)

- ファインチューニングと比較して，とても少ない学習データで高い受容率になったことは，高いポテンシャルを感じさせた.
- 例示の数よりも包括性が重要
- 疑似コードが複雑であると，出力メッセージも長文となった.
- 臨床試験でよく用いられる自然さが欠けたり，実行するたびに結果が異なる.

＊Code LlamaおよびLlama 2: いずれもMeta社が2023年に公開した大規模言語モデル．Code LlamaはLlama 2をベースにプログラムの生成について追加訓練が実施されている.
** Mistral: Mistral AI社が2023年に公開した大規模言語モデル.

クエリ-文作成のまとめ

- **ファインチューニングアプローチ**は，
  - 実際によく使用されるクエリメッセージで訓練させることから，生成されるメッセージの表現も自然な感じのものが比較的多い．十分な量の学習データがあれば人間が介在することを前提に実用できる可能性もあるかもしれないと思われた．
  - 特に正しくクエリメッセージを生成できることがある程度確認できているパターンであれば，安定して正しいクエリメッセージを自然な表現で生成できるのではないかと思われた．
  - 一方，本アプローチは，学習データの量，バランスおよび質がモデルの性能に大きく影響するため，実際の業務で活用するためには，十分な量の偏りのない，一定のルールで作られた学習データのデータベースのメンテナンスが必要になると思われた．
- **大規模言語モデルアプローチ**は，
  - 言語理解能力を持つ大規模な言語モデルを使用して論理チェック仕様からクエリメッセージを生成することから，疑似コードとクエリメッセージの組み合わせを学習する必要もなく，ある程度の品質のクエリメッセージを作成できるようである．
  - 一方，疑似コードを機械的にテキストに置き換えているようで，言語として自然さを欠くメッセージとなってしまうことも少なくない．また同じ疑似コードからクエリメッセージを作成する場合でも，時に微妙に異なるメッセージを生成しているような印象があり実行するたびに結果が異なっていたため，安定性の点ではフ ァインチューニングアプローチに軍配があがるのではないかと思われた．

- 2つのアプローチを通して，学習データの用意の可否や，メッセージをどこまで自然な表現にしたいかといった，ユーザー側の要求を踏まえてアプローチを選択していく必要があることが認識された．

ユースケース

- **コーディング**: CRF（Case Report Form，症例報告書）に報告した用語（有害事象，併用薬等）を標準的な用語に結び付ける.
- **データレビューとクエリ文作成**: CRFデータの不整合を検出し，自動的にクエリ文を生成するプロセスを導入する.

- **併用薬剤**の使用目的の妥当性の判断.

併用薬剤の使用目的の妥当性の判断

- 本業務は，主に当該薬剤の添付文書を閲覧して効能または効果に記載されている文書を読み，CRFに記載された疾患や症状に対して有効かどうかを判定するというプロセスで行われていることが多い．しかしながら，効能または効果に記載されている文書に基づいて使用目的の妥当性を判断することは難しい．
  - 医薬品の有する効能または効果が複数の症状や疾患にわたること
  - 添付文書に記載される表現が製造版販売者によって異なる
  - CRFにおいて薬剤名やその使用目的はフリーテキストで収集されており，特に使用目的は医学辞書コーディングの対象とされていないことも多い

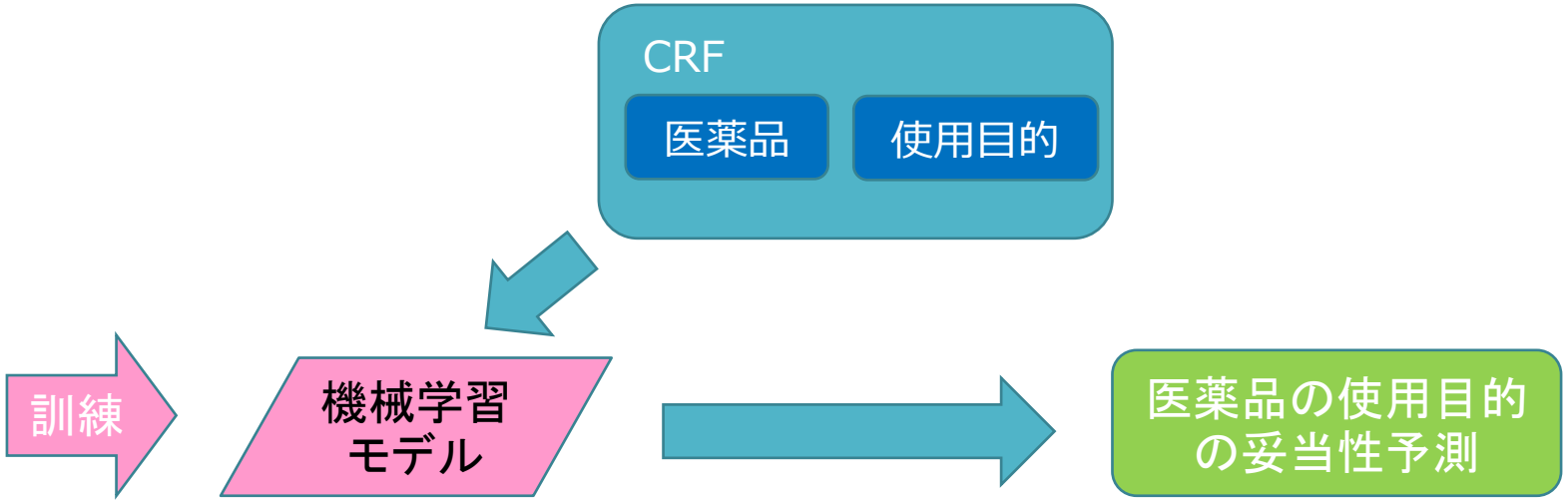
- 多岐にわたる添付文書の表現と，フリー記載で取集されるCRFの情報を照らし合わせて判定する必要があるため，機械的に実施することが困難．
- 企業ではAIを用いてチェックしている事例もある．
- ファインチューニングと大規模言語モデルでアプローチで試みた．

製薬協のTFメンバで試してみた　－併用薬剤の使用目的の妥当性の判断　-

ファインチューニング アプローチ

学習データ

一般名	効能または効果
医薬品 A	アレルギー性鼻炎
医薬品 A	じんましん
医薬品 A	アレルギー性結膜炎
医薬品 B	胃潰瘍
	….

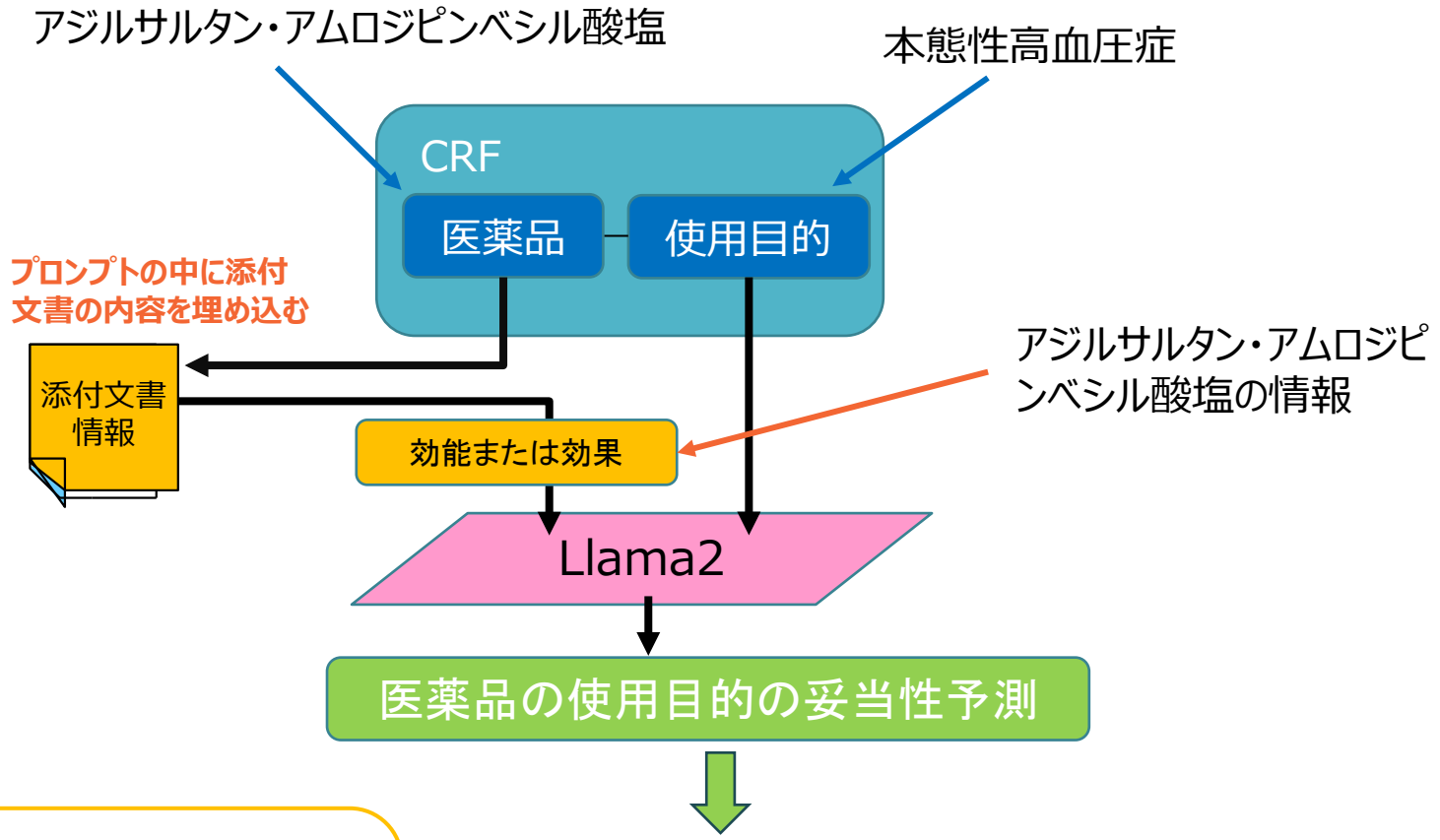


- 「●●に伴う××」といった特定の疾患（●●）に伴う症状（×　×）に対する適応を持つ薬剤の場合は，適応にない疾患（●●）まで抽出されてしまうことがある．そのため，正確な学習データを作成するために非常に時間を要する．
- 添付文書の効能または効果に記載される症状や疾患は，MedDRA 等で一般化された用語ではなく，重症度やその条件によって表記が異なることがある．効果的な学習を行うためには，同様の効能を同じ表現に統一することが望ましかったが，これは効能または効果に記載された各症状や疾患に対して医学辞書コーディングと同等の作業を行う必要があった．
- CRFで収集される疾患に対しても同様に，効果的な判定を行うためには収集された使用目的に対して上記と同等の作業を行うことが望まいが，そのような作業は実務面での煩雑さを考慮すると取り入れるハードルが高かった．

製薬協のTFメンバで試してみた　－併用薬剤の使用目的の妥当性の判断　-

大規模言語 アプローチ

成分名	区分	効能または効果
アジルサルタン・アムロジピンベシル酸塩	内用薬	高血圧症
クロルフェニラミンマレイン酸塩	内用薬	じん麻疹，血管運動性浮腫，枯草熱，・・・・
クロルフェニラミンマレイン酸塩	内用薬	蕁麻疹，血管運動性浮腫，枯草熱，・・・・
クロルフェニラミンマレイン酸塩	注射薬	じん麻疹，枯草熱，皮膚疾患に伴う痒（・・・，・・・・

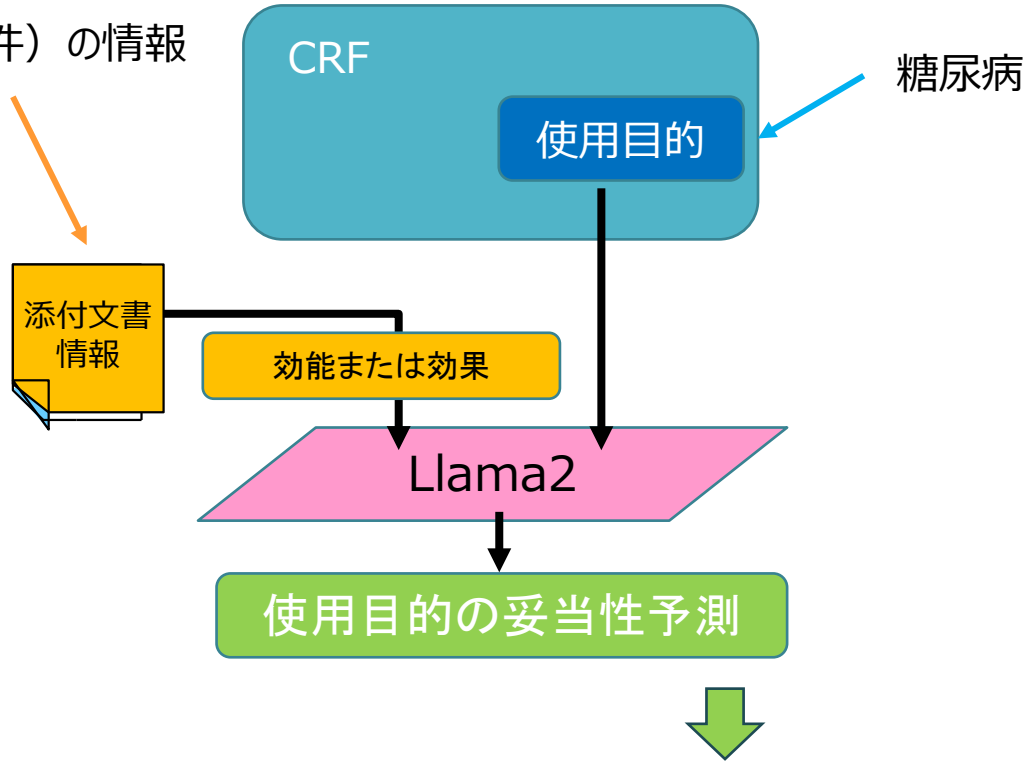


- 学習用のデータが不要になる．
- 承認されている薬剤の一般名，およびその効能または効果の一覧表の作成が必要．
- 国ごとの承認内容に基づいた対応も可能となる汎用性を持ったアプローチとなりえる．

製薬協のTFメンバで試してみた　－併用薬剤の使用目的の妥当性の判断　-

大規模言語 アプローチ

薬剤の一般名（1,819件）の情報
（2023年10月時点）



- 一定の精度は見込めたが，あたかも正しい答えかのように回答してしまう ハルシネーションも認められたため，実務に取り入れる上ではこれらの制御方法も併せて検討が必要.
- 「糖尿病」という単語が含まれているにもかかわらず「効果なし」と判定しているケースが見受けられた.
- 糖尿病の治療に直接的な適応を持たない薬剤に対して「糖尿病に併発する疾患の治療に使用する薬剤であるため効果あり」と予測する，添付文書以外の情報を利用して判定をされているようなケースも見受けられた.

		AIによる予測	
		効果なし	効果あり
添付文書上の記載	効果なし	1,736	21
	効果あり	3	59

併用薬剤の使用目的の妥当性の判断のまとめ

- これまでは困難と思われてきた帳票による点検の自動化も可能な範囲であることが確認された.
- 古典的なプログラムのように確実性のある結果は得られないが、それはある意味で人が行う点検にも同様のことが言える. 不確実なものをどのようにプロセスに組み込んでいくのか，よりリスクを意識したアプローチを念頭に置いて業務の自動化を検討していく必要がある.
- 今回得られている結果は自由に使用可能な大規模言語モデルで実施したものであり，GPT-4に代表される高性能な大規模言語モデルを使用し，プロンプトを練りこめばこれ以上の結果が得られることは想像に難くない.
- 人がミスをすることと同様にAI も完璧ではないことも事実であり，どこでどのように採用していくかはリスクに応じて検討することが必要である.

データマネジメントにおけるAIの活用　- まとめ -

- 今回の BERT および T5 によるファインチューニングアプローチにおいては,ライセンス権および各社の社内機密情報などの観点から使用できるデータは限定的ではあったものの，「疾患名の辞書コーディング」および「論理チェック仕様からクエリー文の作成」では，かなり高い精度であった.
- 同様のモデルを構築する際に各社で蓄積されている社内のデータおよびライセンス権が制限なく利用できるのであれば，より高い精度の AI モデルの構築が期待できる可能性がある.
- 今回使用したLlama2などの LLM （Large Language Models，大規模言語モデル）は学習データを用意することもなく，ある程度の精度の結果を出すことができた. 今後も同様な LLM の発展が期待でき，またノーコード／ローコードによる AI 開発がより身近なものになることで，様々な場面でこれらのモデルを活用できる可能性があることが示唆された.

データマネジメントにおけるAIの活用　- まとめ （続き）-

- AI を実際の業務で活用するためには，利用する状況に応じてRisk-based Approachをとる必要があり，判断や制御に人間の介在が必須である.
- 一方で AI は日々進化しており，AI に関わる様々な問題を理解した上でリスクを適切に管理しながらこれを活用し，それによって業務を改革していくことが今後のデータマネージャーに求められることであろうと思われた.
- AI に関する知識以上に重要なことは，マニュアル作業に依存している非効率な現状に対する問題意識とそれに対する解決策を熟考することである．このような問題意識があれば，AI はデータマネジメント業務に革新的な変化を生み出す未来をもたらすものと思われた.