

カルテ情報から 研究用データベース構築するAI

Seagaia meeting 2023 (2023/5/19, 20)

AICRO株式会社 西畑 俊樹

AICRO

西畑 俊樹

京都大学大学院情報学研究科



リクルート(2008～2016)



きりんカルテ(2016～2021)



AICRO (2021～)

カルテの非構造化
データ(フリー記載)から
重要な情報を抽出して
研究用データベースを
構築するためのAI開発

Seagaia meeting 2019

AIによるカルテ情報を 構造化する医療 言語処理

その後の変遷

きりんカルテ

2018年 クラウド型電子カルテ「きりんカルテ」における、カルテ情報からAIが重要な医療情報だけを抽出、構造化する医療言語処理エンジン「きりんカルテDX」を開発

2019年 Seagaia meeting 2019で発表

2021年 きりんカルテをPHCに事業譲渡

AICRO

2021年 治験・臨床研究における、診療情報の収集・抽出に際して自然言語処理により治験・臨床研究を支援するアイデアを構想し当社設立

2023年 現在、大学病院との共同研究を実施
医療情報匿名化プログラムと構造化AIプログラムを開発中

AICROの紹介

社名	AICRO株式会社 ※ CROが行っているモニタリング業務をAIで自動化できないかというところからAI-CROと名付ける
設立	2021年8月2日
本社	東京都千代田区一番町6 相模屋本社ビル7F
ボードメンバ	山口 太一（代表取締役 CEO） ←きりんカルテのCEO 西畑 俊樹（代表取締役 CTO） ←きりんカルテのCOO兼CTO
事業	<ul style="list-style-type: none">• 自社プロダクト開発• AIを用いたプロダクトやプロトタイプの開発• 主に医療系の会社やプロダクト・開発のコンサルティング• 医療情報システムの開発受託

今日の参加メンバ

**代表取締役 CEO : 山口 太一**

早稲田大学法学部卒。株式会社リクルート入社。その後、コンサルティング会社にて新規事業開発、事業再生案件に多数従事。上場バイオベンチャー入社、執行役員就任。子会社として少額短期保険株式会社を創業、代表就任。きりんカルテシステム株式会社入社、代表取締役就任。(カルテ事業譲渡後、退任)

2019年、きりんカルテシステム子会社のルフトメディカルケア(医療特化派遣業)を総合メディカルに売却

2021年、クラウド型電子カルテ事業をPHC(オンプレ型電子カルテのトップシェア企業)に事業譲渡

**代表取締役 CTO : 西畑 俊樹**

京都大学大学院情報学研究科修了。専門分野は機械学習。元IPA未踏事業のクリエイター。Web系の開発を中心に自らサービスの企画開発をした後、リクルート入社。

人材事業でレコメンドアルゴリズムの開発などプロダクト開発に従事。きりんカルテシステム株式会社入社、取締役就任。

きりんカルテのビジネスと開発を率い、日本医師会ORCA管理機構が提供するレセプトエンジンに連携する電子カルテNO.1としてクラウド型電子カルテ市場拡大をけん引。

**AI Engineer : 渡邊 雅俊**

2008年からエンジニアとしてキャリアをスタートし、様々な業界のシステム開発を経験。2016年からAI領域に専念し、画像認識エンジンの開発や医療データ活用に取り組む。2018年より構造化カルテエンジンの開発にメインエンジニアとして貢献。

現在はAICROでAIエンジニアとして匿名化プロジェクトを進行中。AIの開発には7年以上の経験を持つ。

取り組み紹介 (1/2)

医療情報匿名化

大阪公立大学、医療法人LIGARE、AICRO株式会社、医療現場ノウハウとAI技術を組み合わせ、電子カルテ情報の研究用データベース構築支援を行う共同研究を開始

AICRO株式会社

🕒 2022年12月6日 11時01分

大阪公立大学（運営法人本部：大阪市阿倍野区、学長：辰巳砂昌弘）、医療法人LIGARE 血液内科太田クリニック・心齋橋(所在地：大阪市中央区、：院長太田健介、以下LIGARE) およびAICRO株式会社(本社：東京都千代田区、代表取締役：山口太一・西畑俊樹、以下AICRO)は、多発性骨髄腫患者における医師の自由記述による診療録を用いて自然言語処理技術開発を行い、研究用データベース構築支援を行う共同研究を開始します。

AICRO株式会社

🔍 フォロー

フォローするとどうなる？

4 フォロワー

URL <https://www.aicro.jp/>

業種 情報通信

本社所在地 東京都千代田区一番町6番地
相模屋本社ビル7F

電話番号 -

<https://prtimes.jp/main/html/rd/p/000000001.000113109.html>

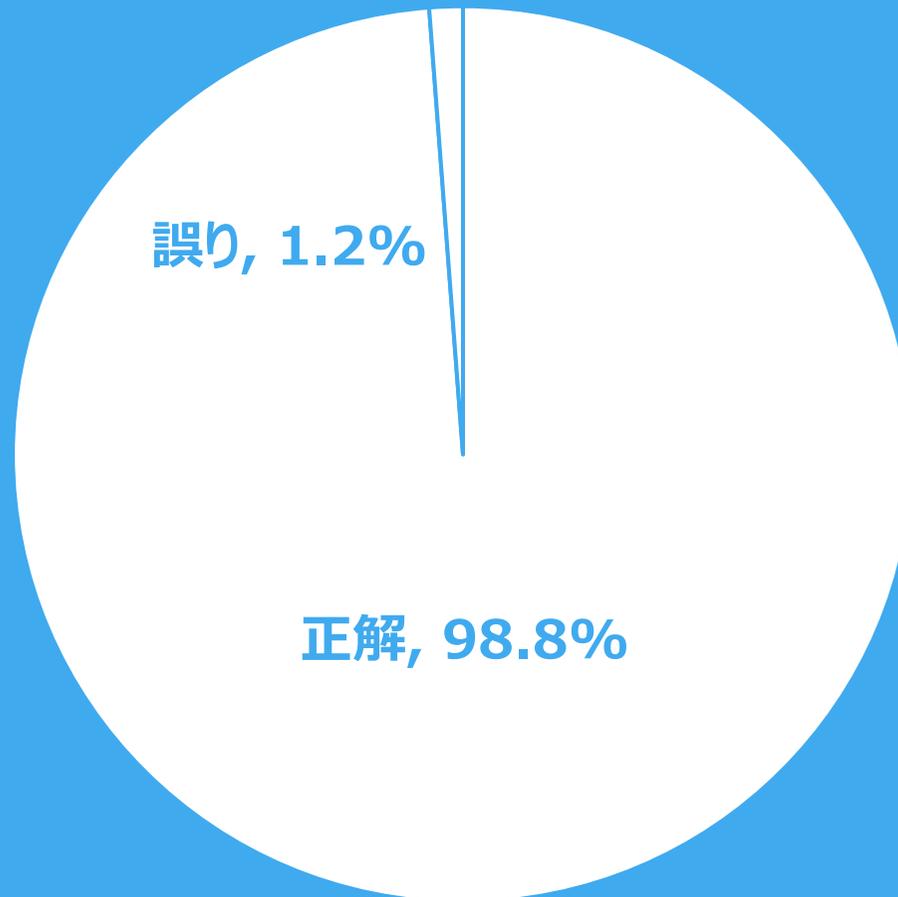
- ・ 病院からデータを外部に出すためには個人情報情報の全削除が必要
- ・ 100%の精度が求められる
- ・ カルテ件数は膨大で手作業でやると日が暮れる

AIと目視チェック&修正を組み合わせ素早くマスキングしたい



※ 特許申請中

医療情報匿名化プログラムのAI精度



※ サンプルカルテ30例のうち、該当個人情報84件中83件を正しく抽出
感度及び特異度のバランスを考えながらAI構築

活用シーン

- 研究で医療情報を外部提供したいが個人情報があるので外に出せない
- 千年カルテ上の非構造化データをもっと活用したい
- ChatGPTなど生成系AIにデータを入れてみたいけど、個人情報があるから試せない



取り組み紹介 (2/2)

構造化AIプログラム開発

2018年

「きりんカルテ」で行っていた構造化カルテでは、テキストで書かれた医療情報（カルテ所見欄）から、「症状」「既往歴」「常用薬」「診断病名」など医学的に重要な情報を抽出し、ICD10等の標準コードに変換して、標準的な情報を抽出。

S: 朝から頭が痛い。鼻水、喉も痛い。咳でる。
熱MAX 38度。下痢嘔吐はなく、食事は取れている。
周囲にインフルエンザなし。孫からうつったのかもしれない。薬希望。
心筋梗塞でカテーテル治療後。
内服：アスピリン、コレステロールの薬(スタチン)

O:
元気そう。
項部硬直なし。
扁桃np、リンパ節触れない。Chest clear

AP:
風邪？
高齢なのでレントゲンだけ取っておく
→肺炎ではなさそう。処方経過観察。



標準症状名	有無
頭痛	+
鼻汁	+
咽頭痛	+
咳	+
発熱	+
下痢	-
嘔吐	-

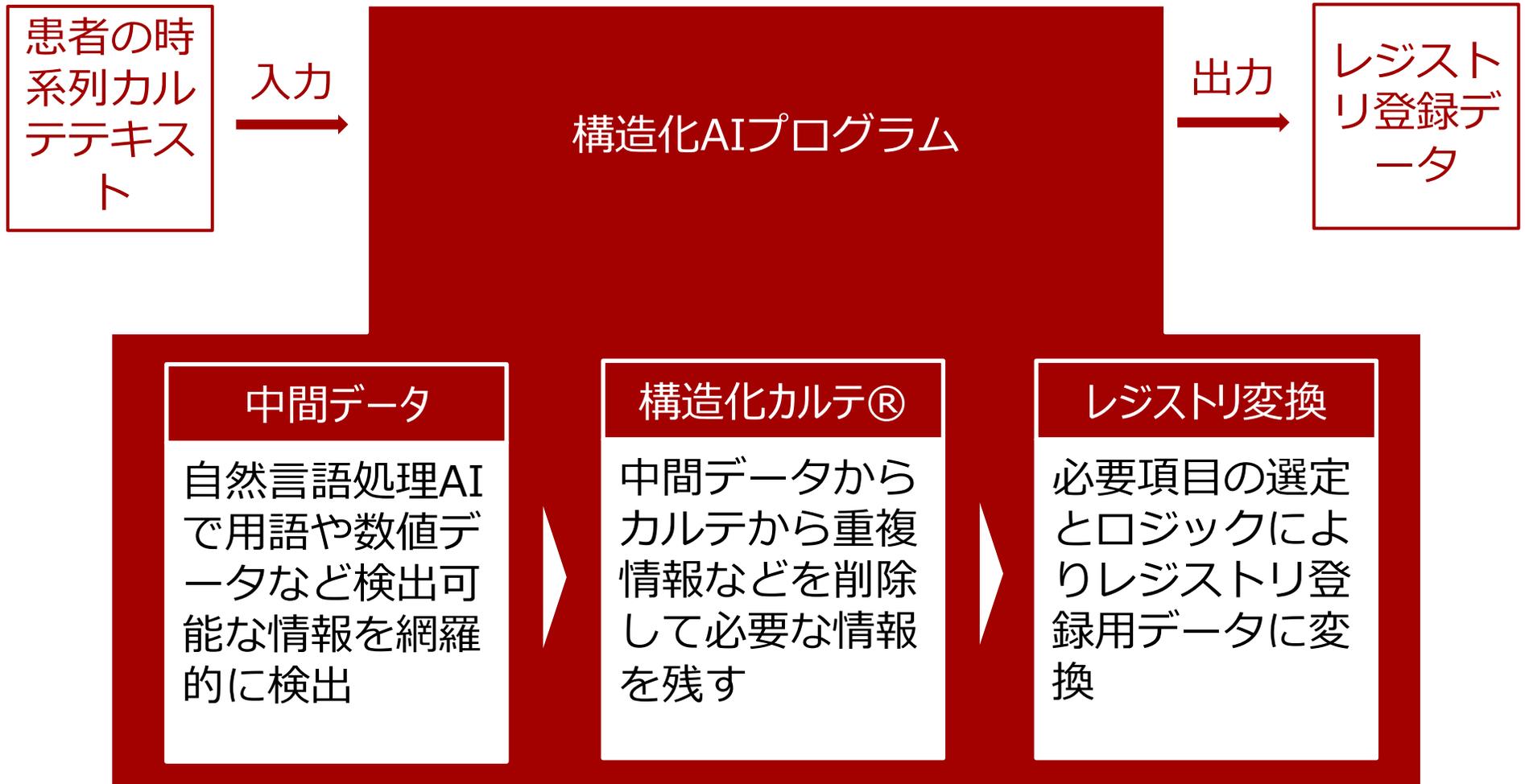
標準化既往歴名	備考
急性心筋梗塞	カテーテル治療後

標準化常用薬名	薬効分類コード
バイアスピリン	3399
スタチン	2189

標準化診断名	備考
急性上気道炎	疑い

進化ポイント

	前回	今回
時系列	単一カルテを対象	時系列のカルテを扱える
用語	標準辞書で対応	主病名等、細かい専門辞書に対応
抽出方法	トピックモデルによるラベル付けと、ロジックによる抽出	Transformerモデルによるラベル付けと、ロジックによる抽出



- 自然な文章ではない。時系列が省略されてる、単語の列挙etc...
- 標準化されていない記載や略称
- 記載箇所によって意味がかわる



プロンプトエンジニアリングを駆使するも、ChatGPT-3.5(最大トークン4,096)、ChatGPT-4(最大トークン32,768)だとルールの提示とfew-shot学習するにはトークン長が足りない...

Model: GPT-4

d あなたは日本の医師です。下記は患者のカルテ記載です。この内容から重要な情報を抽出してください。

""""

LDH f/u

MMK chemo

DM compliance不良

No change

Lt sensation↓ 変わらず
本日はsore throatもある

O

Np

Throat slightly reddish

Neck LAD(-)

Painに対してInjection try

だいぶ良くなったと

Flu s/o心配されており検査→negative

暖かくしてください

次来週

- 自然な文章ではない。時系列が省略されてる、単語の列挙etc...
- 標準化されていない記載や略称
- 記載箇所によって意味がかわる



プロンプトエンジニアリングを駆使するも、ChatGPT-3.5(最大トークン4,096)、ChatGPT-4(最大トークン32,768)だとルールの提示とfew-shot学習するにはトークン長が足りない...

データを集めることが難しい領域だからこそ、特化型AIを開発する意義がある

Model: GPT-4

d あなたは日本の医師です。下記は患者のカルテ記載です。この内容から重要な情報を抽出してください。

.....

LDH f/u
MMK chemo
DM compliance不良

No change
Lt sensation↓ 変わらず
本日はsore throatもある

O
Np
Throat slightly reddish
Neck LAD(-)

Painに対してInjection try

→negative



構造化カルテが完成した未来

- 研究データベースを自動でつくれる。研究者の時間削減
- 製薬の治験データの収集を半自動化できる。創薬のコストの削減
- 重要な情報を抽出できるので、それらのデータを使って、文書生成を自動化できる
- そもそもカルテはメモ帳でいい世界

医療情報匿名化と構造化AI、
非構造化データの利活用に、
ご興味頂けた方はぜひご一緒に取り組
みさせていただきます。



nishihata@aicro.jp



<https://www.facebook.com/toshikisaturday/>