



AI Enginerring Decoded #1

生成AIは科学研究をいかに変え得るか

2024-07-28

fuku株式会社 代表取締役 山田涼太

自己紹介



山田 涼太

X @roy29fuku

f roy29fuku

- 2010年4月 東京大学 農学部 獣医学専修
- 2016年3月 休学届け提出
- 2017年4月 東京大学 工学部 システム創成学科 へ転学部
- 2018年3月 fuku株式会社 創業
- 2019年3月 東京大学 工学部 システム創成学科 卒業

生命科学実験の効率化を入りに
科学 × AIの領域で活動

趣味

ミステリーADV

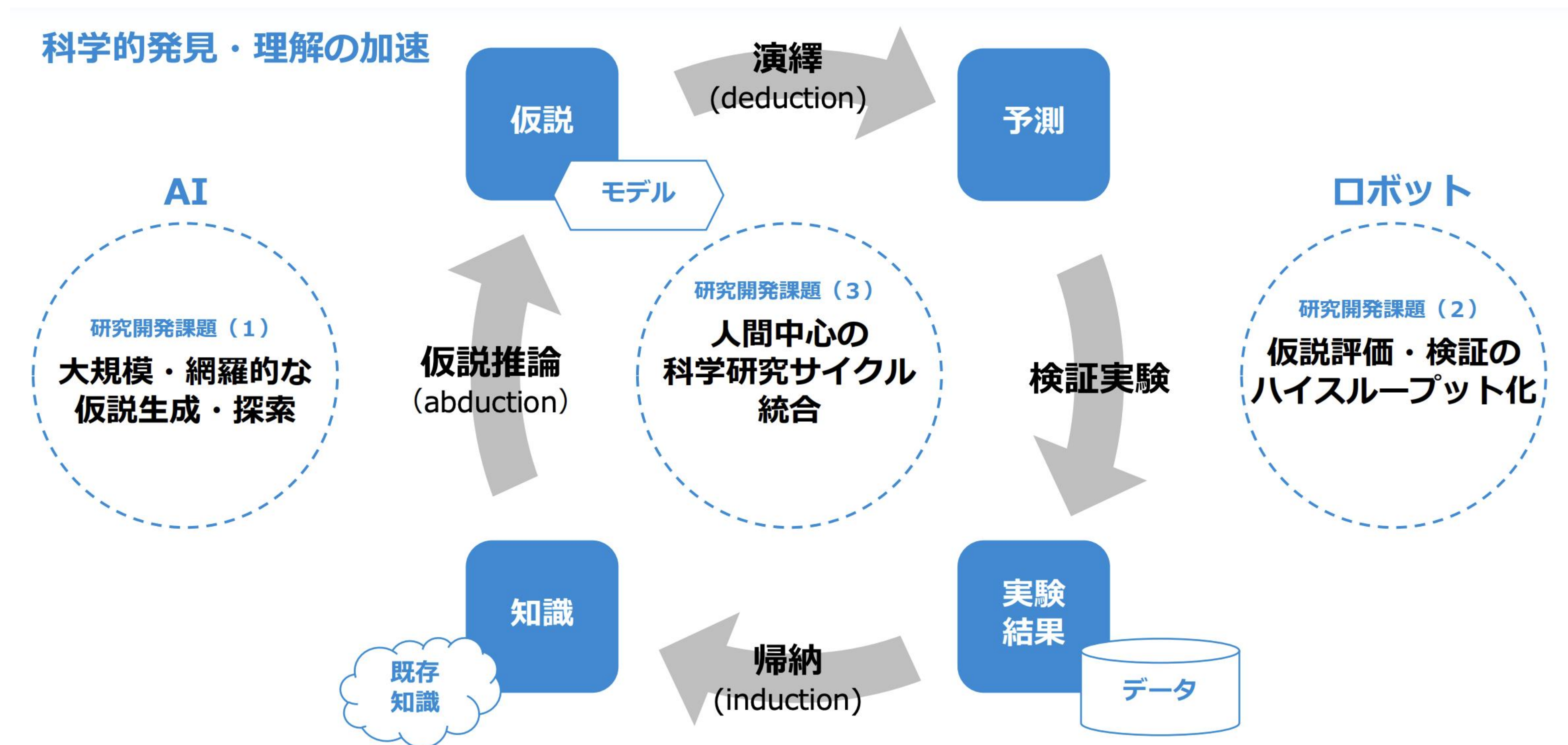
ダンガンロンパ、逆転裁判、極限脱出、レイジングループ

最近面白かったやつ



研究の自動化を目指しています

科学研究のPDCAサイクルにAIやロボットを活用するAI・ロボット駆動科学に取り組んでいます



会社紹介：具体的な取り組み

理研：ロボティックバイオロジーによる生命科学の加速

ロボットによる生命科学実験を実現するためのプロトタイプングラボ。双腕型のロボットが人間と同じ実験機器を利用して人間の代わりに実験を行う。
実験データの自動解析や**実験ログの収集・可視化**に携わっています。



京大：「富岳」で目指すシミュレーション・AI駆動型次世代医療・創薬

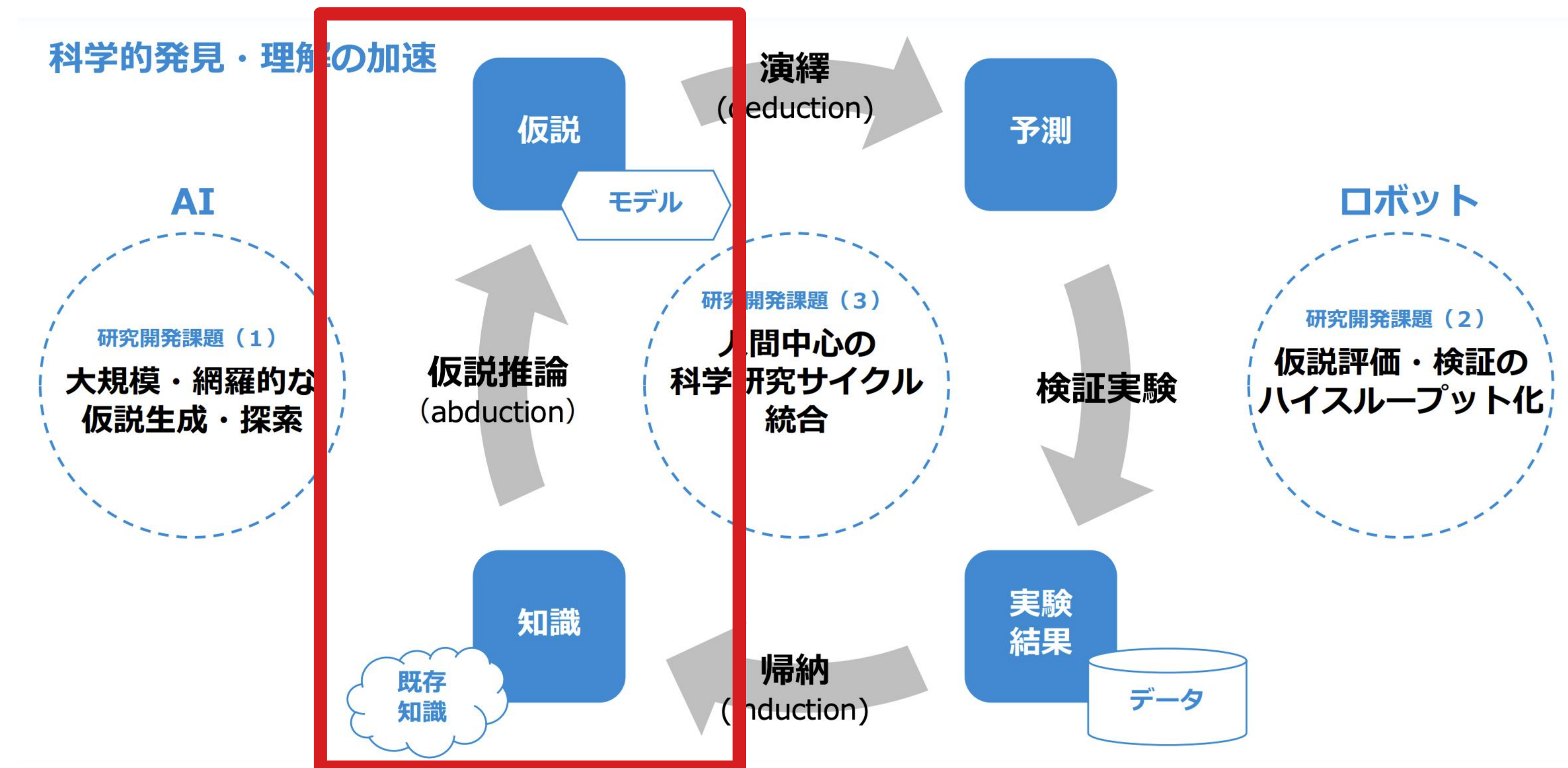
創薬AIプラットフォーム。産官学連携により、ライフサイエンス系公共データベースや製薬企業の実験データを用いて様々なAIモデルを開発する。
公共データの構造化や**パイプライン構築**に携わっています。



本日のテーマ

既存知識から仮説を生成する部分へのAI活用にはフォーカス

生成AI登場以前の取り組み、現在の取り組み、どのように進めているか、何が難しいか



出典：文部科学省 基礎研究振興部会「基盤モデルとAI・ロボット駆動科学」

生成AIの登場で**顧客課題の解決にフォーカス**できる

- ✅ 基盤モデルの性能向上により、少量の正解データさえあればIn-context learningで多くのケースで実現可能性をクイックに検証可能
- ✅ Agentic workflowを採用することで、ユーザーの入力が不十分・不適切な場合をカバーできる可能性がある

👉 既存知識を入力として新たな仮説（リサーチクエスション）の出力を目指す

原著論文

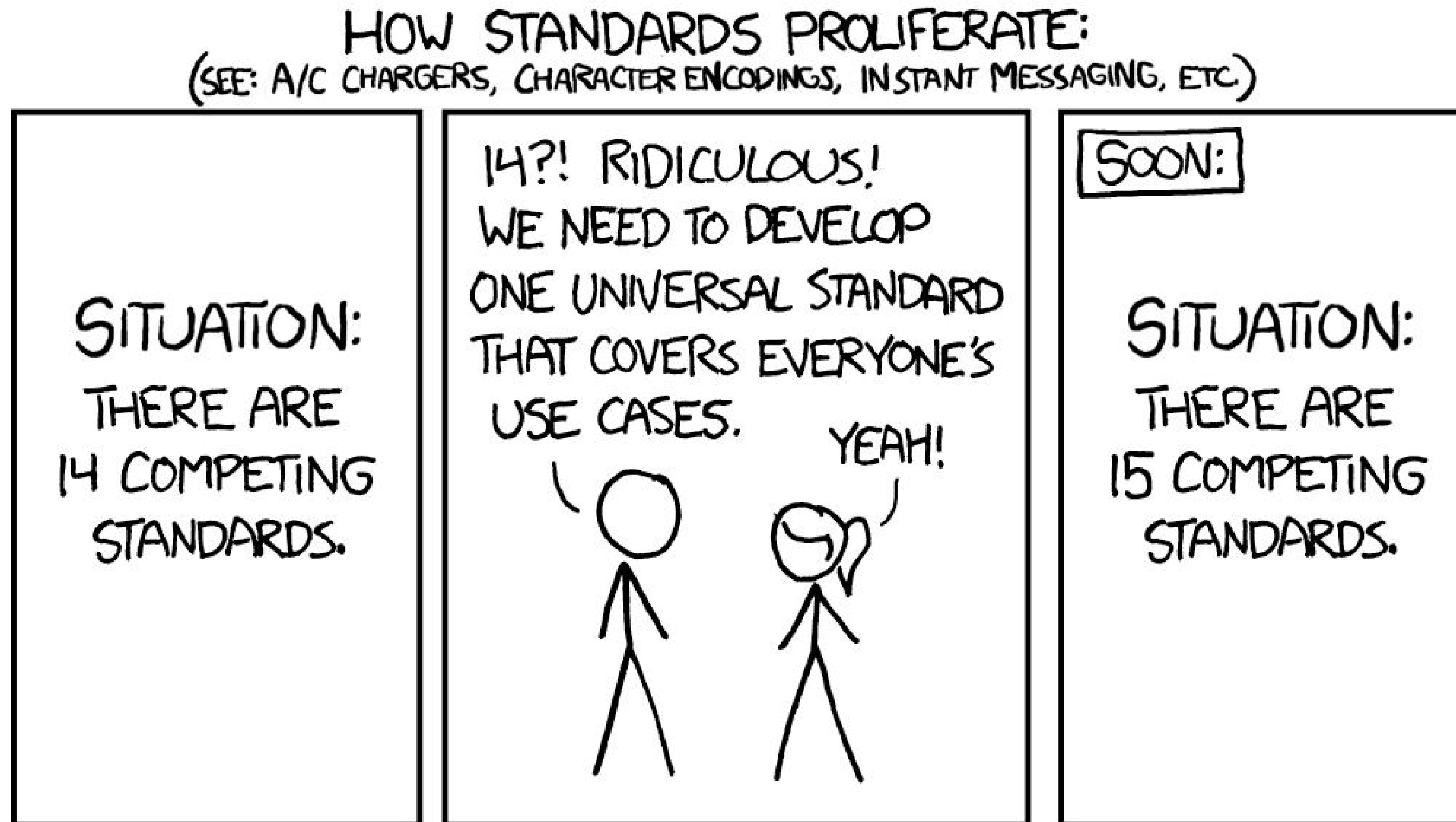
研究の背景、手法、結果、考察などを含む非構造化・半構造化データ
テキストだけでなく、FigureやTableには画像やテーブルがある

公共データベース

特定の関心ごとに関する情報を集約した構造化データ
多くはRDBMSが採用されるが、RDFなどのグラフデータや独自フォーマットのテキストデータの場合もある
ドキュメントが整備されていないことが多い

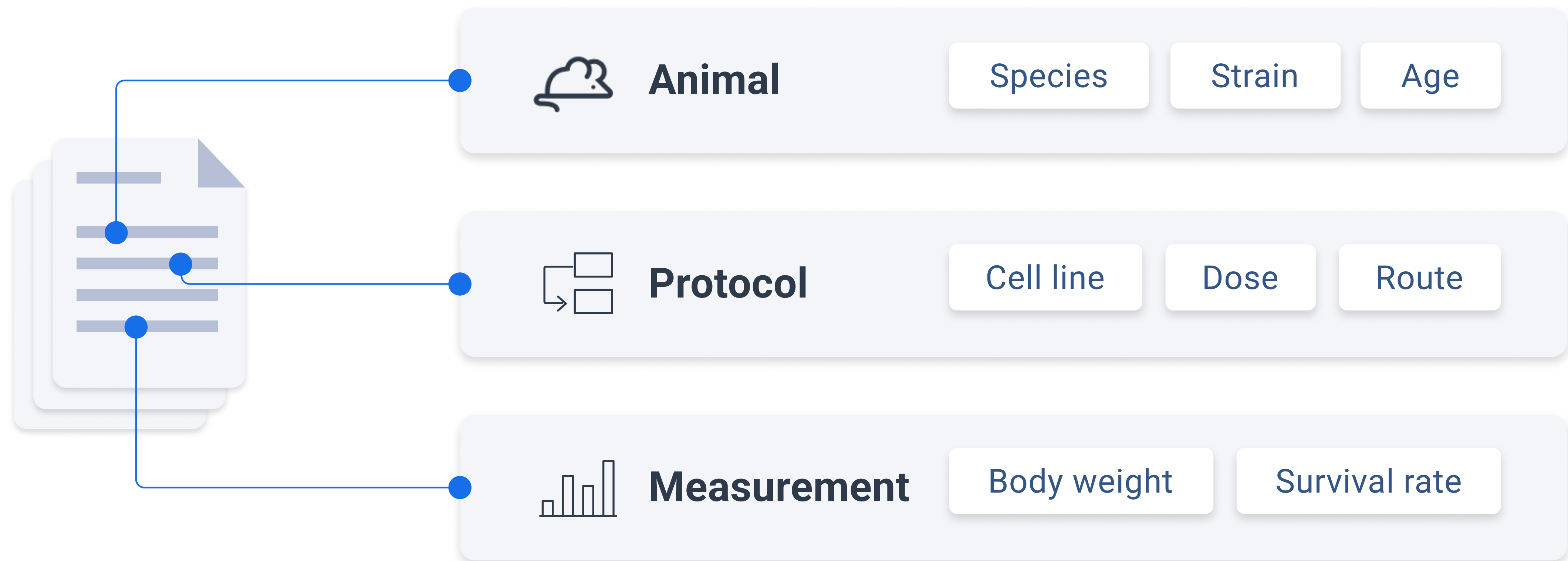
背景

既存知識を効率的に活用するため、構造化や統一フォーマットの導入などの様々な取り組みがされてきているが、未だ解決されていない（というか問題が増えている気さえする...）



Before Generative AI

約5年間、デジタルネイティブな知識の蓄積フォーマットを開発



抽出

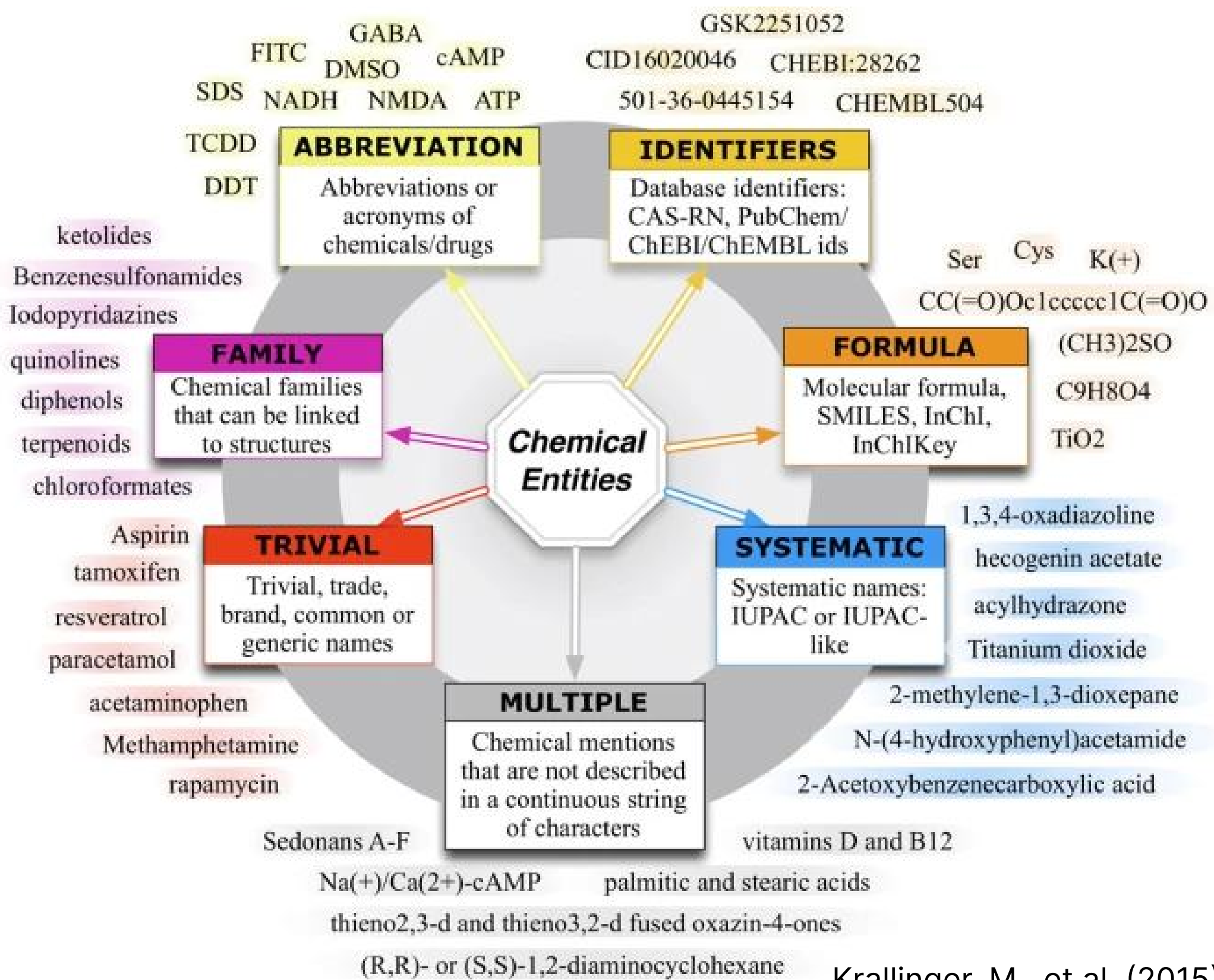
- Named Entity Recognition
- Relation Extraction

To further elaborate on the geographical trends, **North America** **LOC** has procured **more than 50%** **PERCENT** of the global share in **2017** **DATE** and has been leading the regional landscape of **AI** **GPE** in the retail market. The **U.S.** **GPE** has a significant credit in the regional trends with **over 65%** **PERCENT** of investments (including M&As, private equity, and venture capital) in artificial intelligence technology. Additionally, the region is a huge hub for startups in tandem with the presence of tech titans, such as **Google** **ORG**, **IBM** **ORG**, and **Microsoft** **ORG**.

<https://medium.com/@alessandropaticchio/named-entity-recognition-from-scratch-e76b9b3affad>

対応付け

- Semantic Web
- Linked Data



Krallinger, M., et al. (2015).

Before Generative AI

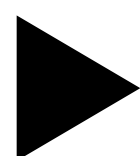
抽出

- Named Entity Recognition
- Relation Extraction

論文などの専門文書から

- 遺伝子名
- 疾患名
- 薬剤名
- タンパク質名

などの名称と関係性を抽出



対応付け

- Semantic Web
- Linked Data

抽出した情報を

- NCBI Gene ID
- MONDO ID
- ChEMBL ID
- UniProt ID

に対応付け

Before Generative AI：課題

最終的に特化型BERTをfine-tuningする手法を採用

└─ 実験手法ごとに1,000~2,000件の論文をアノテーション

💀 ロングテールに対応できない

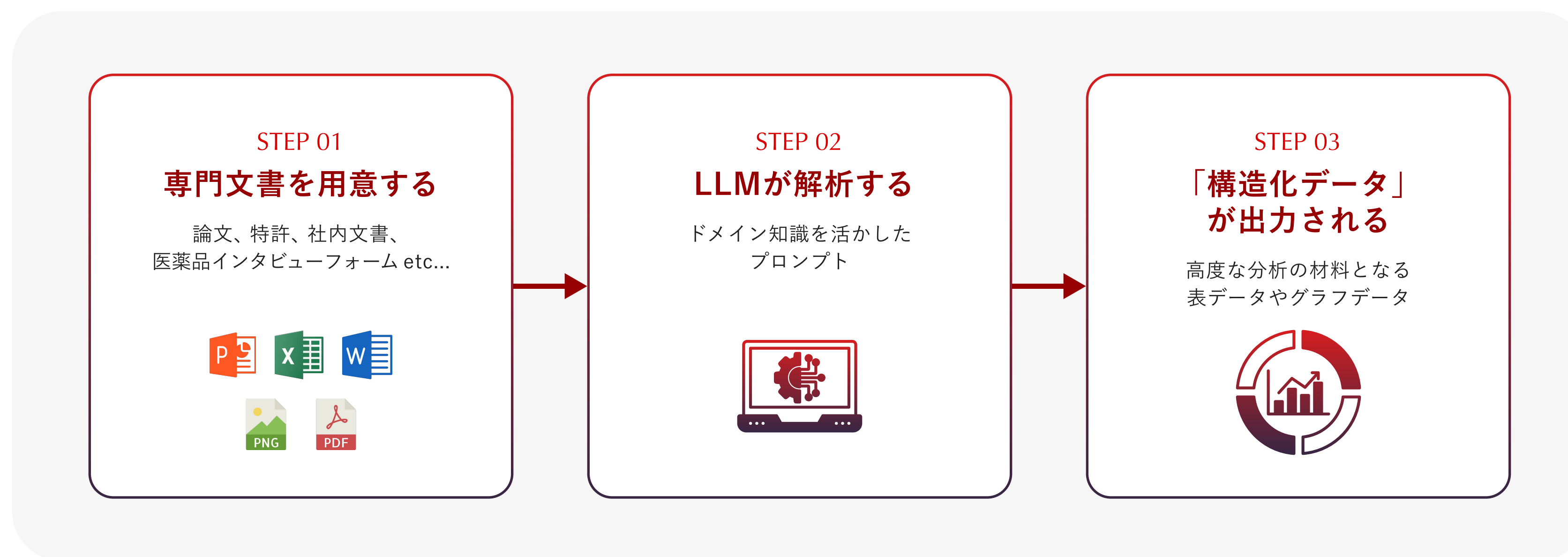
そもそも対象の論文が1,000件程度しかないこともある
十分な学習データを作成できるリソースがあるならAIが必要ない

With Generative AI

生成AIの活用により、顧客課題の解決にフォーカスできるようになってきた

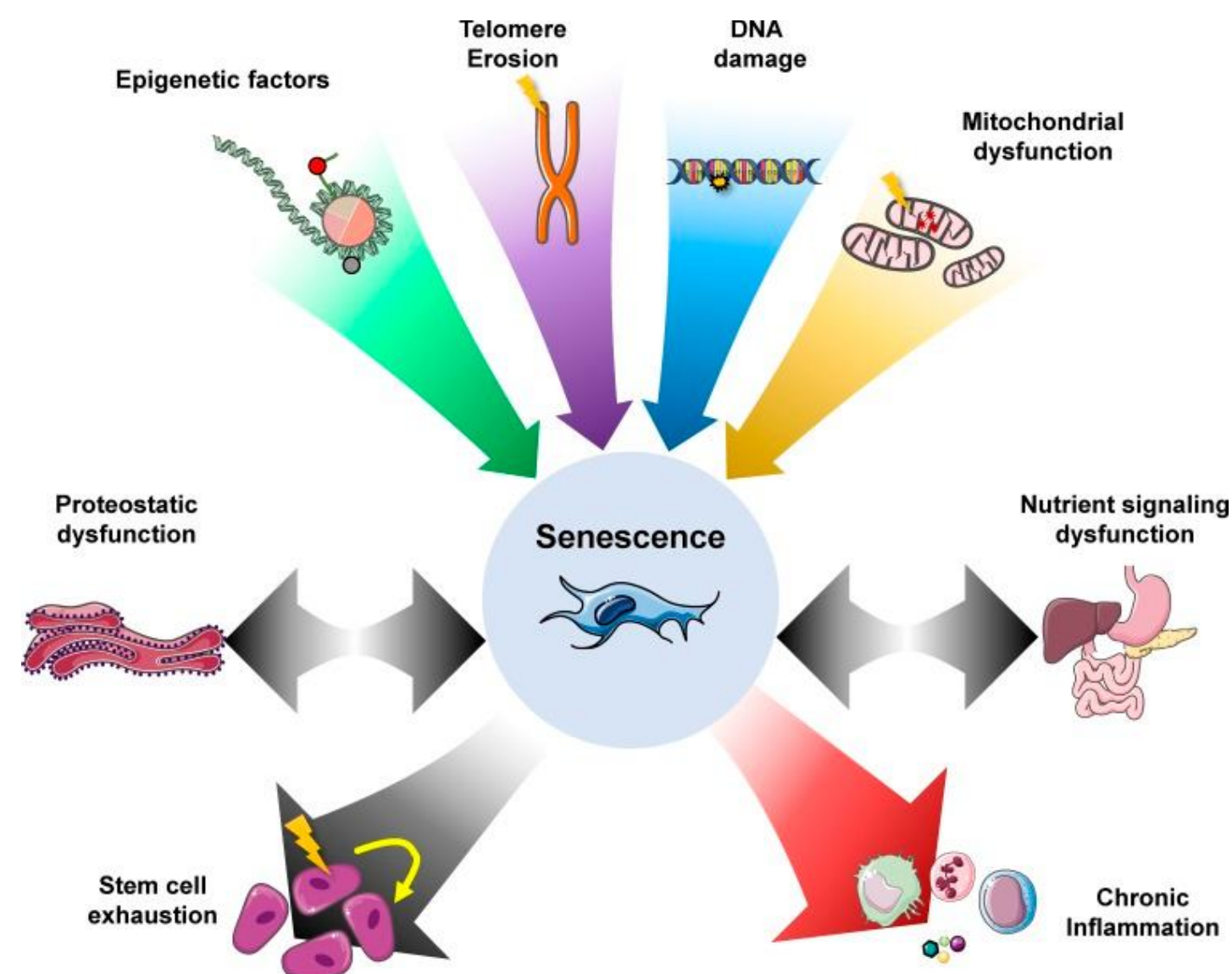
1. 論文からの情報抽出
2. データベースを対象としたRAG
3. Agentic workflow

1. 論文からの情報抽出

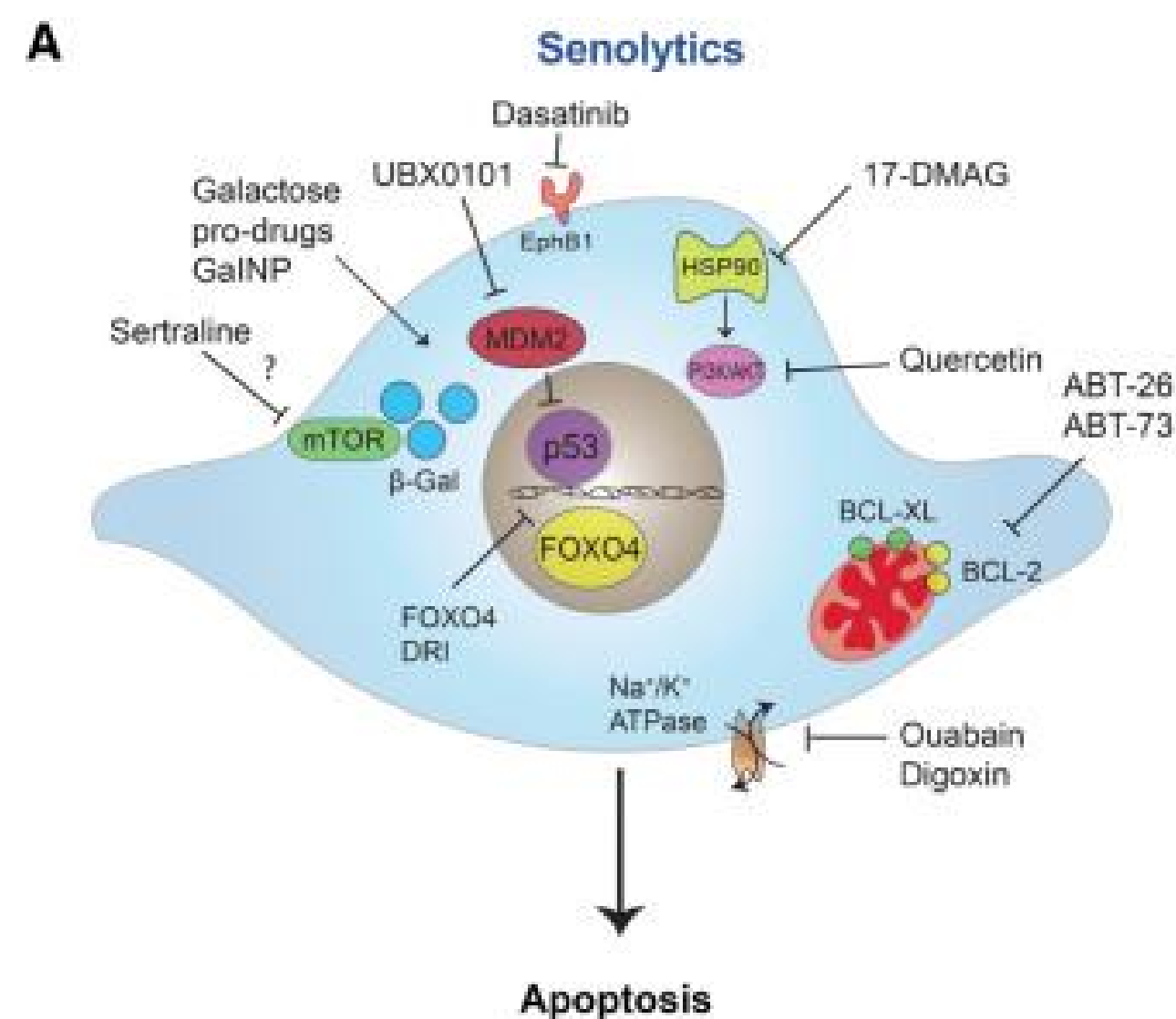


- 【筑波大学】 論文から細胞の形態や機能に関する記述の抽出
- 【DBCCLS】 医薬品インタビューフォームの表データの抽出
- 【理研BRC】 論文から化合物、タンパク質、遺伝子などの生命科学関連の単語の抽出
- 【非公開】 論文のTableデータの抽出

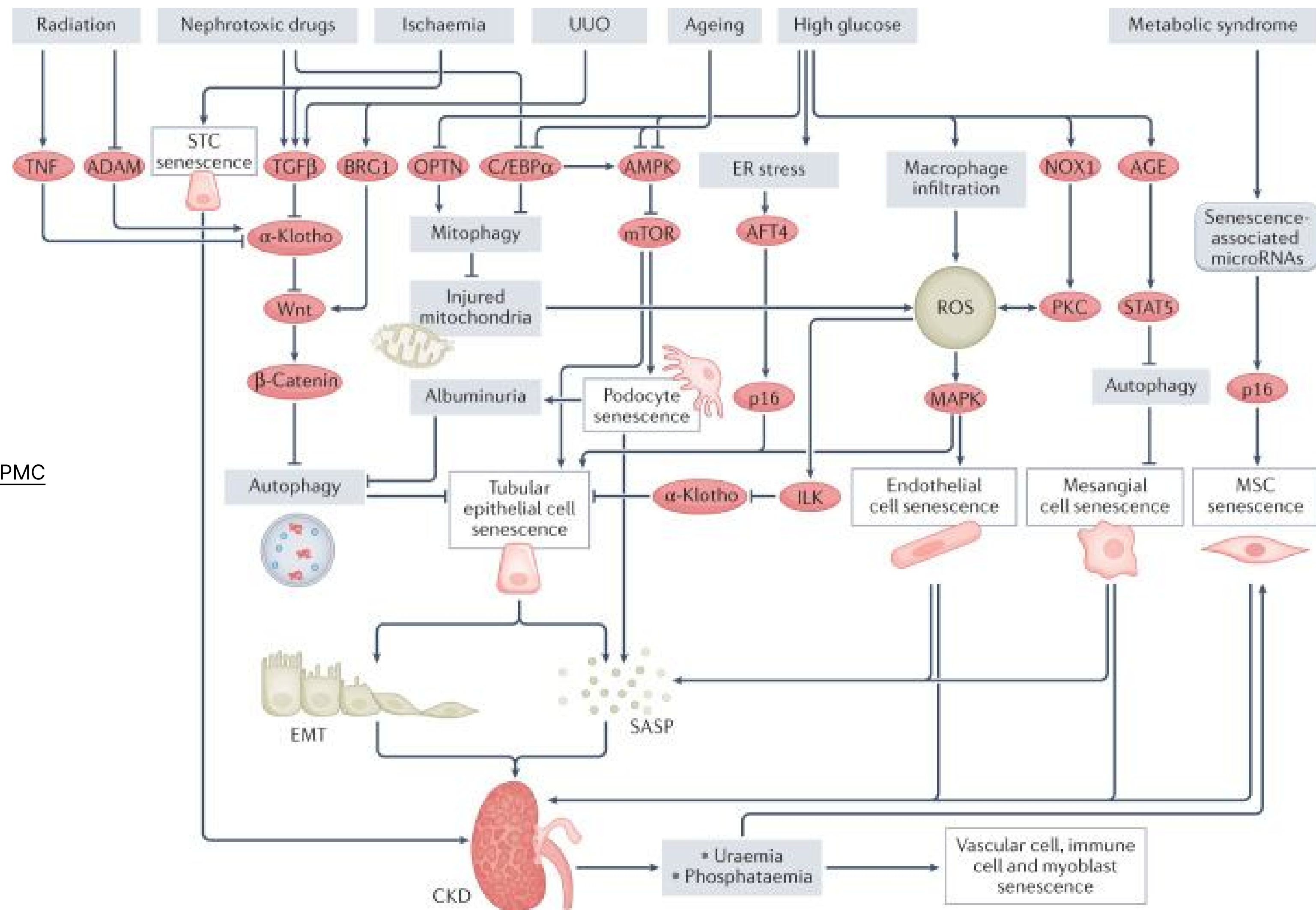
1. 論文からの情報抽出



出典：Senescence and aging: Causes, consequences, and therapeutic avenues - PMC



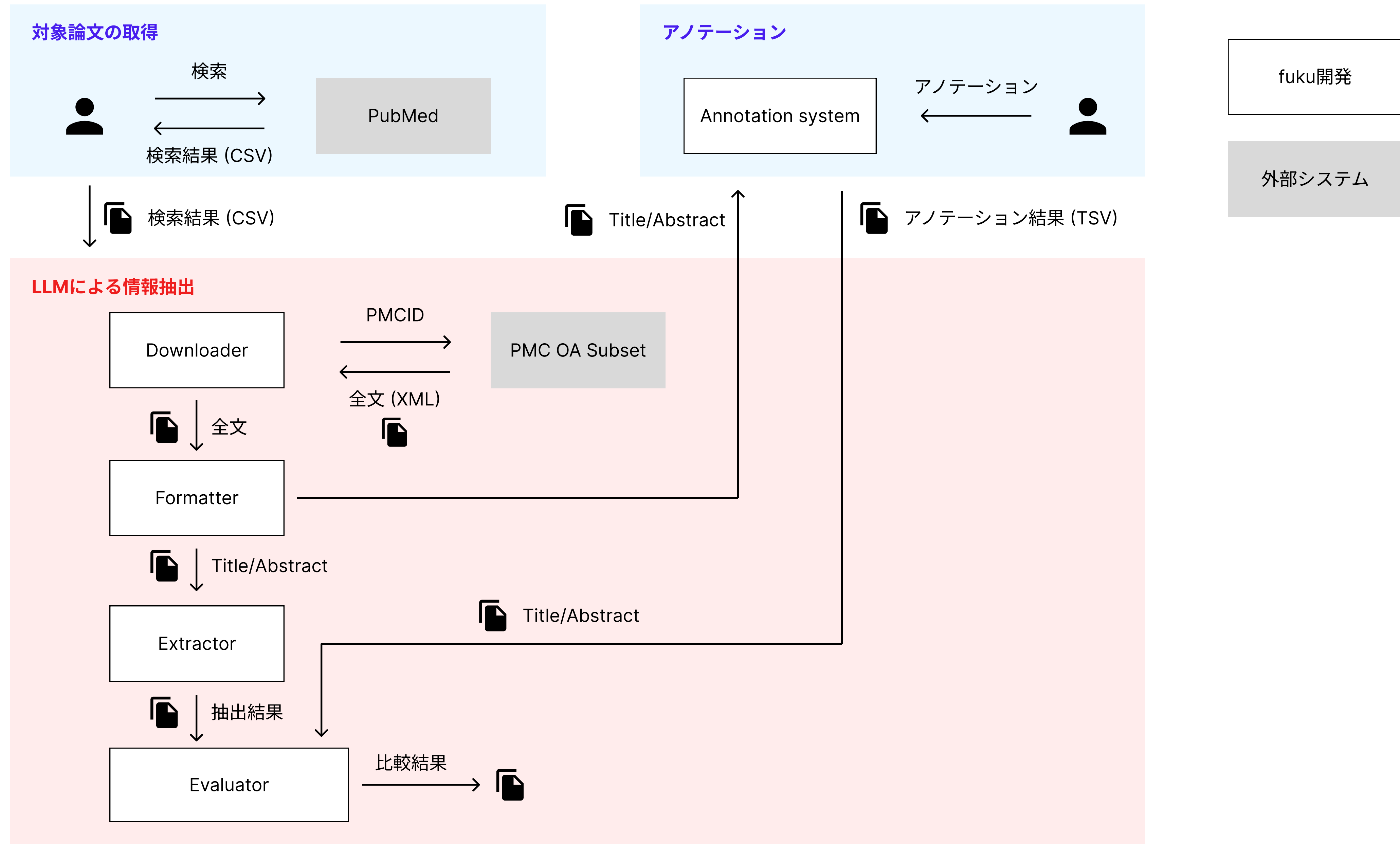
出典：Senescence and the SASP: many therapeutic avenues - PMC



出典：Cellular senescence: the good, the bad and the unknown - PMC



1. 論文からの情報抽出



2. データベースを対象としたRAG



- 【京都大学】創薬AIのための学習データセット生成
- 【農研機構】植物の系譜情報に関する質問応答

2. データベースを対象としたRAG

As-is

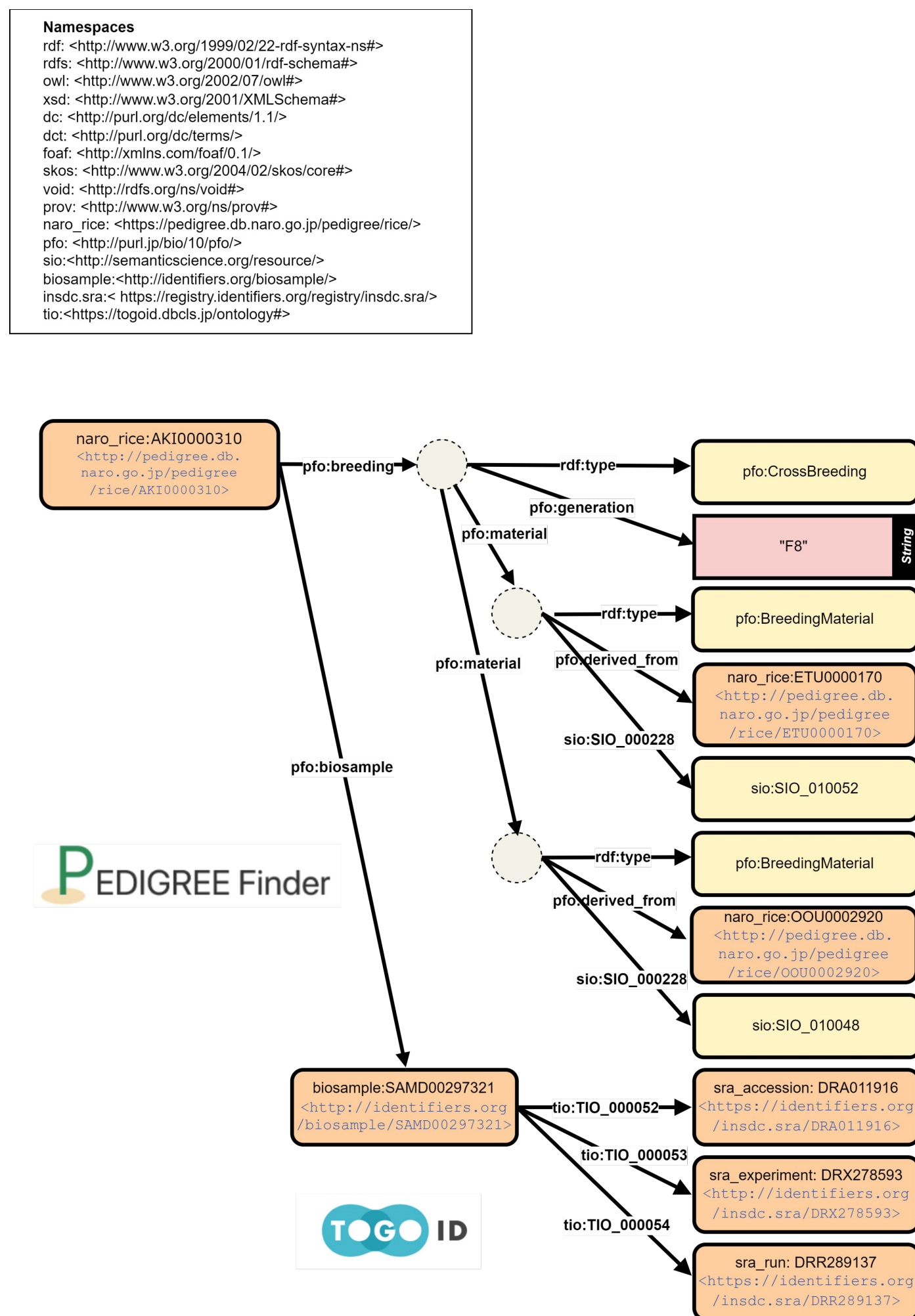
- ライフサイエンス領域では様々なDBが存在
- 使いこなすにはドメインと技術の両方の知識が必要

To-be

- LLMにDBのSchemaを理解することで操作させることが可能
- ユーザーは自然言語で問い合わせをし、LLMがクエリを発行

実績

- 【農研機構】 自然言語でライフサイエンスDBへの問い合わせ
- 【京都大学】 創薬AIの学習データの自動生成



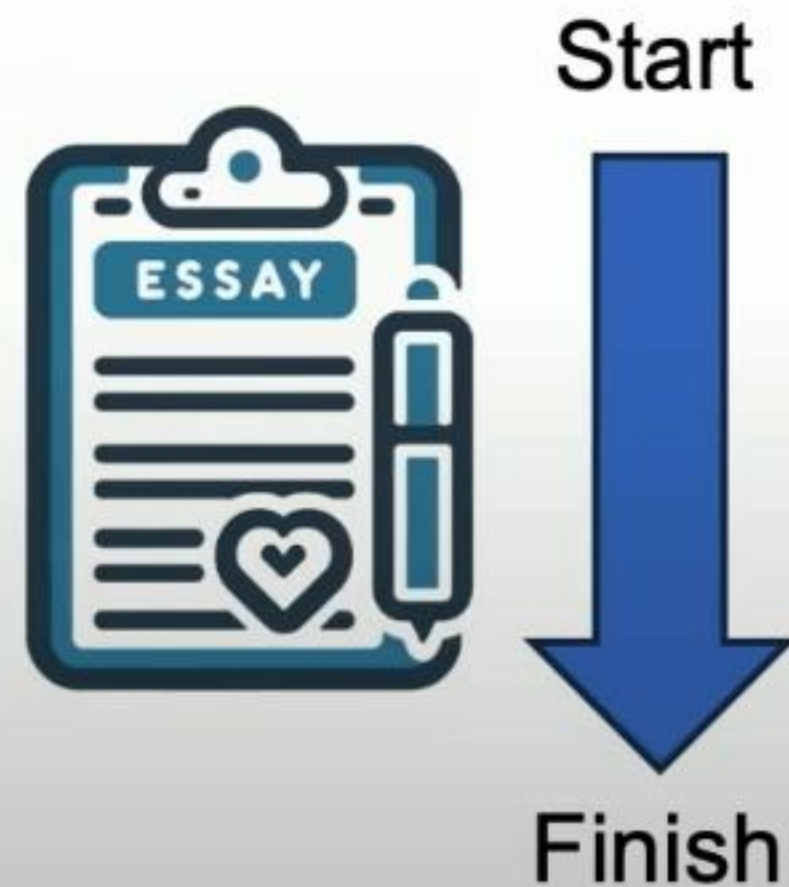
<https://github.com/dbcls/pfo/blob/main/schema/PedigreeFinderSchemaCrossing.png>

3. Agentic workflow

LLM-based agents

Non-agentic workflow (zero-shot):

Please type out an essay on topic X from start to finish in one go, without using backspace.



Agentic workflow:

Write an essay outline on topic X

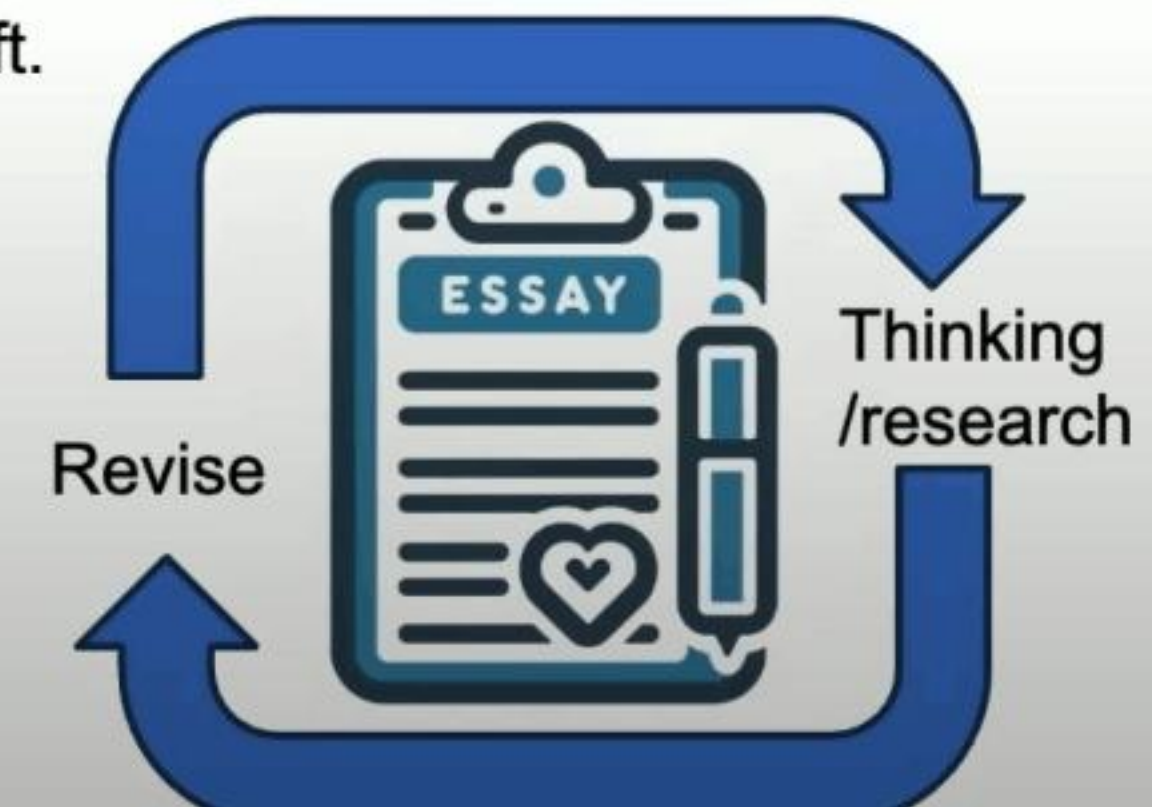
Do you need any web research?

Write a first draft.

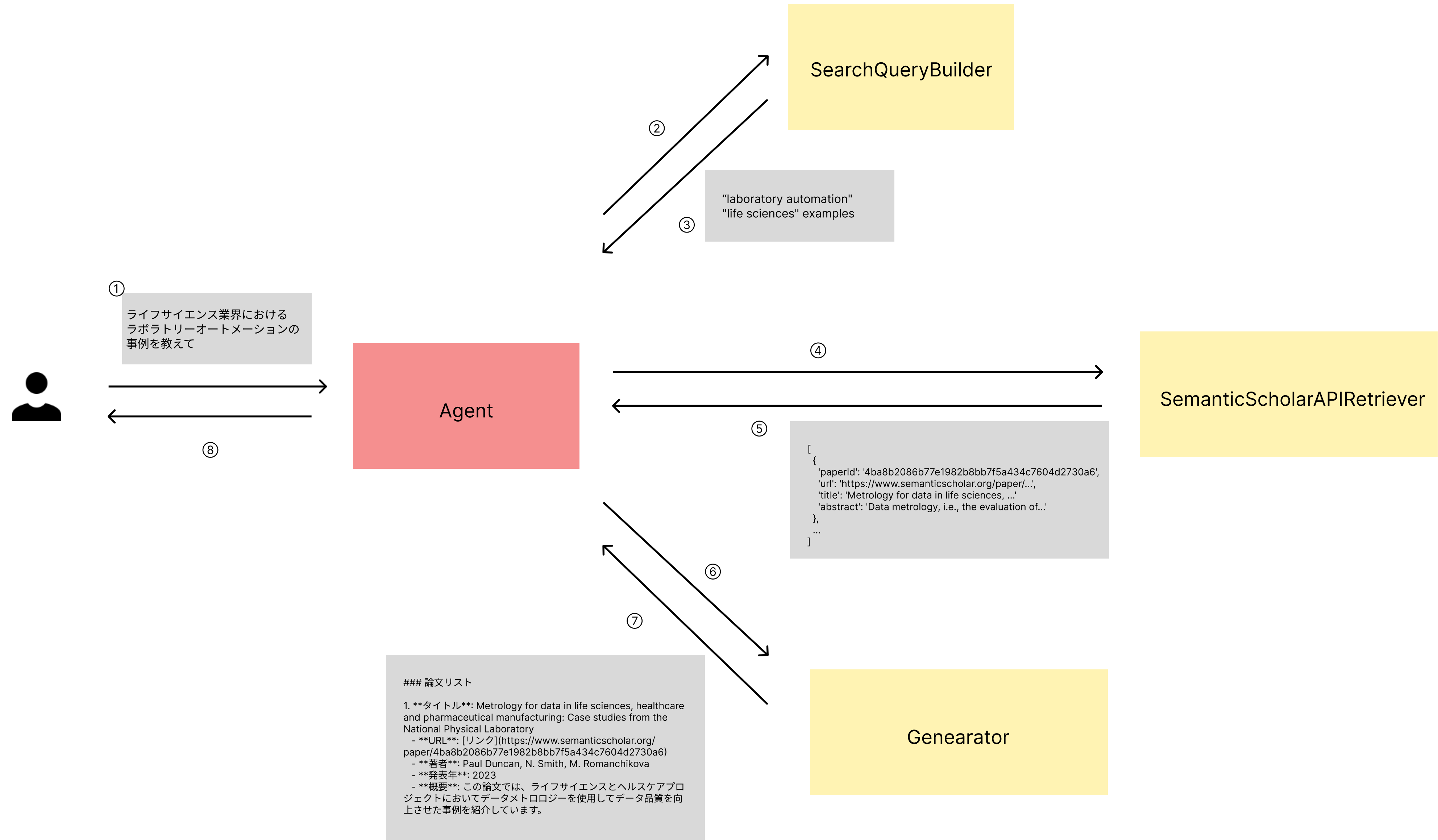
Consider what parts need revision or more research.

Revise your draft.

....



3. Agentic workflow



 We are hiring !

AIロボット駆動科学は今！熱い領域
興味がある方は一緒にお仕事しましょう！

こんな方におすすめ

-  **科学の自動化を実現し人類に貢献したい**
-  **AI・ロボットに科学研究をやらせて何が生まれるか知りたい**