

# 人文学研究の「高度化」と「民主化」に向けた AI活用モデル

— 西洋古典対話システム「Humanitext」の実践と、  
次世代の研究基盤「AI Readyなデータ」の構築 —

---

## 発表者

岩田 直也

名古屋大学 デジタル人文社会科学研究推進センター 准教授  
国立情報学研究所 特任研究員（BOOST 次世代AI人材育成プログラム）



# Humanitextプロジェクトの概要 — 原典と対話するRAGシステム

## システムの核心

100著者・1000作品超から意味的に最適な「原典テキスト」をピンポイント抽出  
すべての回答に原典出典（Source）を必ず紐づけ、研究水準の検証性を確保



## 独自技術「文脈指向翻訳（Context-Oriented Translation）」

原典をそのまま検索せず、AIが理解しやすいよう「主語」や「背景」を補った中間データを生成。  
これにより、低リソース言語（古典語）でも高精度な意味検索を実現。



# AI活用がもたらした成果 — 異分野融合と国際展開

## ➡ 異分野融合・新分野創成

哲学・歴史・文学を横断した調査・分析が可視化・高速化

Antiqua（西洋古典）を核に、アラビア語圏／中世・近世欧州へ歴史横断的に拡張予定

→ 従来の学問領域を超えた、新たな研究視点と方法論の創出

## 🌐 国際性の向上 Japan-led Innovation

欧米ツールの受容から、知の発信へ転換

日本発システムとしてグローバルコミュニティに新たな選択肢を提供

→ 世界に先駆けた人文学研究基盤の構築により、国際的プレゼンスを高める

# 研究にAIを活用し始めたきっかけとハードル



## 意味検索の必要性

キーワード一致ではなく、概念や文脈に基づく最適なテキスト発見

Serendipity（偶然の発見）を支援する研究環境の構築

## ハルシネーション対策

生成AIの「もっともらしい嘘」を防ぐ

原典に基づく事実確認を自動化したいという動機

→ 研究の信頼性を担保しながら、AI技術の利点を最大化



## ハードルと克服

### プログラミングの壁

人文学者にとって技術的実装は大きな障壁

従来は専門エンジニアへの依存が必須

### Vibe Codingによる突破

生成AIによるコーディング支援の活用

「何を作りたいか」の解像度さえ高ければ実装可能

→ AIとの協働による人文学者の技術的エンパワーメント

# 研究者のスキル —「設計者」としての視点

問い合わせ：人文学者はどこまで情報系スキルが必要か？

→ フルスタックができる必要はない

必須スキル：研究者を「設計者（アーキテクト）」として位置づける



## アーキテクト能力

研究課題解決に必要なツール・データ構造の「設計図」を描く力

設計思考



## 検証（Verification）能力

AIやツールの挙動が、自身の「研究手法」として適切か判定する力

批判的評価



## 協働力

他分野の研究者と共同作業を行うための共通言語と指揮能力

学際連携

# データリポジトリの課題 —「AI Ready」なデータとは

## ⚠️ 既存アーカイブの限界

### | TEI/XMLの構造的問題

木構造（Tree）：保存・表示には最適

文脈が階層上位に存在するため、断片化に弱い

AIから見ると「文脈が欠落したデータの羅列」

### | AI活用における欠点

高度な推論が困難

意味的な関連性の把握が不十分

暗黙知が明示化されていない

→ 人間には読めるが、AIには理解しにくいデータ構造

## ✓ AI Readyなデータ

### | 文脈完結型チャンク

断片化しても「誰が・いつ・何を」が自立して理解可能

暗黙知の明示化による再利用性の向上

AIが直接処理できる形式での構造化

### | ナレッジグラフへの再構築

エンティティ（人物・概念）をノードとして定義

解釈（アイロニー・学説）もノード化

多層的に結合された意味のネットワーク（Graph）

→ テキスト間だけでなく、概念・解釈レベルでの相互参照を実現

# なぜドメイン専門家が必要か —「解釈」の構造化

## ■ GraphRAG導入の必然性

通常のRAGでは、文献間に潜む複雑な相互参照（Intertextuality）や背景知識を扱えない  
意味のネットワークとして多層的に結合された構造が必要

## ■ Human-in-the-Loop人文学者の役割

エンティティ同定：異名・婉曲表現の実体結合

例：「アリストンの息子」= プラトン

関係性の意味づけ：引用／批判／皮肉／贊意など「解釈」のデータ化

注釈・先行研究の知見をエッジ属性として実装し、推論の精度と説明可能性を向上

→ 蓄積された学術的知見を構造化できるのは専門家のみ

# AI時代における人文学の「勝ち筋」と基盤強化

## 1. 人文学研究者の「勝ち筋」：コンテキスト戦略

計算資源を競うLLM開発ではなく、「入力データ（コンテキスト）」の質で勝負する。  
独自の高品質なドメイン特化型知識グラフがあれば、汎用モデルであっても世界最高のアウトプットを創出可能。

## 2. 制度設計

### 評価システムの刷新

論文至上主義からの脱却

質の高い「構造化データセット」の構築や公開そのものを、論文と同等の「主要な研究業績」として評価する仕組みが不可欠

### ハイブリッド人材の育成

伝統的な文献学の知見に加え、データ構造（グラフ・オントロジー）の設計知識を併せ持つ「次世代型人文学研究者」の養成が急務

## ▲ 3. 重大なボトルネック

### 人材不足：現状の危機

学部生は多いかもしれないが、博士課程に進み「ドメイン知識」を深める研究者は激減

AI Readyなデータを作るための「専門知識を持つ人材」自体が枯渇しつつある

#### ➡ 予算配分への提言

ドメイン知識を持つ人が減るような予算の振り替えは本末転倒。「基礎研究（ドメイン人材）への支援」を厚くした上で、プラスアルファとして「AIデータ構築」に予算を投じる必要がある。

# 結び：まとめと行動提案

01

技術

Context-Oriented Translationにより、低リソース言語（古典語）でも高精度な意味検索・対話を実現。

02

役割

研究者の役割は、AIの出力に対する「設計（Architecture）」と「検証（Verification）」へシフトする。

03

データ

TEI（保存用・木構造）から、AIが推論可能なKnowledge Graph（活用用・グラフ構造）への転換が不可欠。

04

勝機

LLMのモデル性能ではなく、入力する「コンテキスト（ドメイン知識）の質」こそが人文学の武器となる。

## 行動提案



コンテキスト戦略



AI Readyなデータ構築



ドメイン人材の評価革新