

足場は動き続ける

——AI との 8 ヶ月の協働から見た人文学のこれから

岩井優士 (2025B006)

東京大学大学院 人文社会系研究科

1. はじめに

基盤モデルの性能向上とツールの進化に振り回され続けた 8 ヶ月だった。

2025 年の夏、日本財団 HUMAI プログラムに採択されたとき、私の研究計画は近代詩のデジタルアーカイブ構築だった。研究資料の電子化と、AI による典拠・影響関係ネットワークの可視化。近現代詩の研究者で、高校国語科の非常勤講師を 5 年務めている人間としては、まっとうな計画だったと思う。このプロジェクト自体は頓挫したわけではなく、新しいモデルやサービスの登場に合わせてワークフローを組み直しながら、むしろ効率化が進んでいる。ただ、その過程で蓄積したノウハウが思わぬ方向に広がった。Claude Code で何ができるか試す中で手書き答案の採点支援 AI を作り、学術論文の知見を動画に変換するパイプラインを組み、教育支援プラットフォームの設計にまで手を出している。プログラミングの正規教育を受けたことがなく、コードを自力で書く能力はほぼない人間が、である。

面白い。面白いのだが、厄介でもある。できることが拡張されていくのは間違いなく望ましいことだが、先月作ったものが今月にはもっと簡単に、もっと良く作り直してしまう。足場が常に動いている。

本稿は、この振り回された 8 ヶ月の実践を報告しつつ、一つの問いに向き合うものだ。基盤モデルの進化が専門知のあり方を根底から揺さぶる中で、人文学の専門家にはいったい何が残るのか。近現代詩のデジタルアーカイブ構築、採点支援 AI の開発、学術知の動画化。これらの実践を経て見えてきたことを、学術論文ではなくエッセイの形で、実践の手触りごと伝えたい。

2. デジタルアーカイブの構築：動き続ける最適解

2.1. やろうとしたこと

日本財団 HUMAI プログラムへの応募時、私の研究計画の柱のひとつは「近現代詩のデジタルアーカイブ構築」だった。青空文庫が著作権切れの文学作品を、国文学研究資料館が歴史的典籍をカバーする一方で、近現代詩の研究に必要な資料の多くはデジタル化が進んでいない。詩論や同人誌、批評の全集といった資料は、書誌情報こそデータベースで検索できても、本文が電子化されていないものが大半で、全文検索はもちろん、テキストを横断的に参照すること自体が難しい。この状況を打開するために、以下のようなパイプラインを構想した。

紙の全集・評論集をスキャンし、OCR（光学文字認識）でテキストデータに変換する。OCRの出力には誤読や不自然な改行が大量に含まれるため、それをクリーニングする。クリーニングしたテキストを適切な長さに分割し、大規模言語モデル（LLM）による校正にかける。校正済みのテキストに書誌メタデータ（著者、作品名、初出誌、初出年、収録書、出版年など）を付与し、構造化されたデータとしてクラウドの検索基盤に登録する。スキャンから検索可能なアーカイブまで、7つのステージからなるパイプラインである。

このパイプラインの中で特に工夫を要したのが、LLMによる校正のプロセスだった。学術テキストの校正は、一般的な文章の校正とは異なる。旧仮名遣いや歴史的な文体を「現代語に直す」のではなく、原文の表記を最大限に尊重しつつ、OCRのノイズだけを除去しなければならない。そこで私は9つの校正ルールからなるフレームワークを設計した。脚注番号のノイズ除去、年月日表記の正規化、詩の引用部分の改行の維持、そして最も重要なルールとして「文脈異常の検出とマーキング」——AIが判断に迷う箇所を無理に修正せず、[要確認]タグを付けて人間の最終チェックに委ねる仕組みである。

このフレームワークを用いて、近現代詩関連の資料をデジタル化し、メタデータを付与する作業を進めた。メタデータの設計にもこだわった。詩や批評の作品は、雑誌への初出と後の全集への収録とで異なる文脈を持つ。そこで初出情報（掲載誌名、掲載年）と収録書情報（出版社、出版年、ページ番号）を分離した2層構造のスキーマを設計し、1つの作品が持つ複数の出版コンテキストを記録できるようにした。

2.2. 何が起きたか

ところが、このパイプラインを構成する個々の技術的アプローチは、次々と最適解が変わっていくことになる。

最大の転機は Gemini 3 の登場だった。このモデルの画像処理能力は、それまでの OCR サービス（Google Cloud Vision OCR など）を質的に超えていた。単に文字を読み取る精度が高いだけではない。ヘッダーやフッター、ページ番号、脚注番号といった要素を「どういう機能を果たす文字列か」として認識し、本文の読み取りに際してスルーするか含めるかを文脈に応じて制御でき

る。私が校正ルールの中で苦心して設計した脚注ノイズの除去や文脈異常の検出を、モデルが画像の段階で処理してしまうのである。

OCR だけではない。コンテキストウィンドウ（一度にモデルに入力できるテキスト量）の急速な拡大により、テキストを数値に変換して類似度で検索する RAG（検索拡張生成）と呼ばれる従来型の手法の必要性が低下した。長大なテキストをまるごとモデルに読み込ませて直接推論させることが現実的になったからだ。ただし、外部の知識ソースを参照して生成するという広義の RAG パターン自体が消えたわけではなく、エージェントが自律的に検索戦略を組み立てる形に姿を変えて残っている。私のプロジェクトも、この変化の中でまったく別のアプローチに移行することになる。

そして何より、パイプラインの「構築」という行為自体が変質した。Claude Code や Codex のような AI コーディングエージェントの登場により、7つのステージからなるパイプラインを一から設計し、スクリプトを書き、テストし、運用するという作業の多くを、自然言語での指示によって短時間で実現できるようになった。数ヶ月かけて設計した仕組みと同等か、それ以上のものを、新しい技術を使えばレビューを重ねながら数日で構築できてしまう。そういう状況が現実になりつつあった。

2.3. 引き継がれたもの

では、やり方が変わり続ける中で何が残るのか。

第一に、電子化されていない専門資料をデジタルデータとして整備する作業そのものは、どんなに AI の能力が向上しても前提として必要であり続ける。AI モデルが未学習の、特定分野の研究者しかアクセスしない資料——近現代詩の全集や同人誌のバックナンバーなど——は、物理的にスキャンしてデータ化しない限り、どんな高性能なモデルにも読ませることができない。手法が古くなっても、「前提を作る仕事」は残る。

第二に、パイプラインの設計経験は、そのまま次の構築に引き継がれた。現在、私は『日本近代文学』と『昭和文学研究』の学術論文約 200 巻分を手始めに、新たな OCR パイプラインを構築している。将来的にはこのデータベースを研究者が利用できるツールとして公開することを目指しており、今後さらに大規模な拡張を計画している。なお、対象とする学術論文の多くは著作権保護期間内にあるが、日本の著作権法における情報解析のための複製に関する規定（30 条の 4）を踏まえつつ、利用形態の法的な整理は今後の課題として残っている。

それはそれとして、パイプライン自体の構築は進んでいる。Gemini 3 によるページ単位のテキスト化、品質フラグによる自動検証（反復エラーや OCR アーティファクトの検出）、目次からの論文境界の自動推定、論文ごとのファイル分割と YAML メタデータの付与、そしてカタログの自動生成。6つのステージからなるこのパイプラインは、最初のプロジェクトで得た知見なしには設計できなかったものだ。

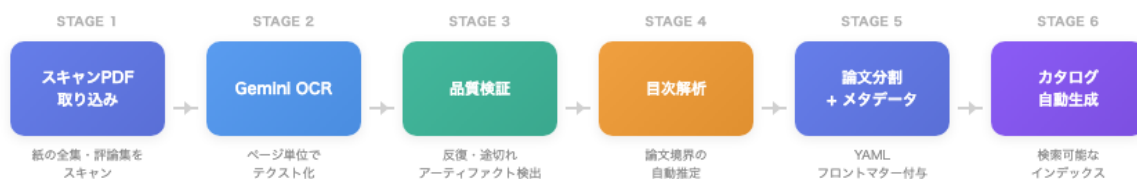


図 1: OCR パイプラインのワークフロー — スキャン PDF からカタログ生成までの 6 ステージ

第三に、そしてこれが最も重要な変化だが、「AI に資料を検索させる」という行為そのもののあり方が変わりつつある。最初のプロジェクトでは前述の RAG を前提にしていた。しかし、コンテキストウィンドウの拡大と AI エージェントのツール使用能力の向上により、まったく別のアプローチが現実的になっている。

新しいパイプラインでは、論文ごとに YAML フロントマター（機械にも人間にも読める形式で書誌情報を記述する書き方）で、雑誌名、巻号、著者、キーワード、対象作品、時代区分、論旨要約などの構造化メタデータを付与し、それを集約したカタログを Markdown のテーブルとして保持している。AI エージェントがこのカタログをテキスト検索し、関連する論文を特定し、必要なファイルを読み込んで推論する、という流れだ。

このメタデータの設計には、「エージェントにいかにか効率よく検索させるか」という明確な意図がある。たとえばキーワードと対象作品を分離したのは、「田村隆一について論じた論文」と「田村隆一の商品そのもの」を混同せずに検索できるようにするためだ。時代区分を独立したフィールドにしたのは、「1950 年代の詩」のように特定の時期に関する横断的な問いに応えるためである。つまり、想定される問いの型からメタデータの構造を逆算している。ベクトル埋め込みのような計算コストは不要で、スキーマの変更も柔軟にでき、何より検索の根拠が人間にも透明に見える。私の規模のコーパスでは、この方法で十分に機能している。

Li et al. (2025) がロングコンテキストと RAG を体系的に比較した研究では、タスクによって両者の優劣が入れ替わるものの、単純なチャンクベースの RAG に対してはロングコンテキストが優位な場面が多いことが示されている。また Singh et al. (2025) が「エージェント型 RAG」として整理しているように、従来の「検索してから生成する」という静的なパイプラインから、エージェントが自律的に検索戦略を組み立てる動的なアプローチへの移行が、この分野では急速に進んでいる。私のプロジェクトは、この潮流のなかで「エージェントが効率よく探索できるようにデータを整備する」という方向に舵を切ったものだ。

個々のアプローチがより良い選択肢に置き換わり続けるのは、個別の技術選択の問題ではなく、AI と資料の関係そのものが変わりつつあるという構造的な変化の一端だ。

3. 採点支援 AI の開発

3.1. きっかけ

採点支援 AI の開発は、Claude Code で何ができるかを試す中から生まれた。

国語科の教員にとって、記述式答案の採点は最も時間のかかる業務のひとつである。一問一問、生徒の記述を読み、模範解答と照合し、部分点を判断し、コメントを書く。定期試験の時期には数百枚の答案がまとまって押し寄せる。非常勤の私ですらその負荷は身に染みているが、専任教員はさらに校務分掌や部活動指導を抱えており、負担は一層重い。専任の先生方がこの負荷を嘆いているのは以前から聞いていた。

AI コーディングエージェントの能力を探る中で、この課題に AI を使えるのではないかと思いついた。HUMAI への応募時点から「AI×教育」を研究テーマのひとつに掲げていたこともあり、最もニーズの明確な課題から着手することにした。

3.2. 根幹システムと継続的な開発

開発には Claude Code (Anthropic 社が提供する AI コーディングエージェント) を用いた。前述のとおりコードは書けないので、開発の実態は、自然言語で「こういうものが欲しい」と伝え、AI が生成したコードを実行し、結果を確認してフィードバックを返す、というサイクルの反復である。自分がやっていたのは、何を作るべきかの判断と、出力の批判的な検証だった。

根幹のシステム、つまり答案 PDF のアップロードから AI による OCR と仮採点、教員による確認・修正までのワークフローは、2 日間で動作するところまでたどり着いた。しかし、実際の教育現場に投入できる品質に上げるには、そこからさらに数ヶ月にわたる継続的な改良が必要である。現在、開発はイテレーション 36 を数える。これは「2 日で何でも作れる」という話ではない。2 日で核ができるからこそ、残りの時間を、現場の教員が本当に必要としているものは何かという設計の本質に集中できるということだ。



図 2: 採点支援 AI のワークフロー — 入力から出力までの 6 ステップ。教員の判断が必要な箇所を明示する設計

3.3. 教員の感覚が規定する技術設計

このシステムの設計を貫いているのは、「完全自動化ではなく、人間の判断が必要な箇所を可視化する」という思想だ。AI が判断に迷う答案にはフラグを立て、何について教員の確認が必要なのかを具体的なコメントとともに提示する。採点の最終判断は常に教員の手にある。5 年間国語を教えてきた感覚が、この設計思想の土台にある。

たとえば、空間プロンプティングと呼んでいる 2 段階の OCR 手法がある。学校の答案用紙のレイアウトは千差万別で、統一フォーマットは存在しない。そこでまず AI に答案画像の全体を分析させて各設問の解答領域を特定させ、次にその領域を高解像度で切り出してテキストを読み取らせる。固定テンプレートに依存する設計では現場では使えない。答案用紙の多様さを知っていたから、最初からこの方式にした。

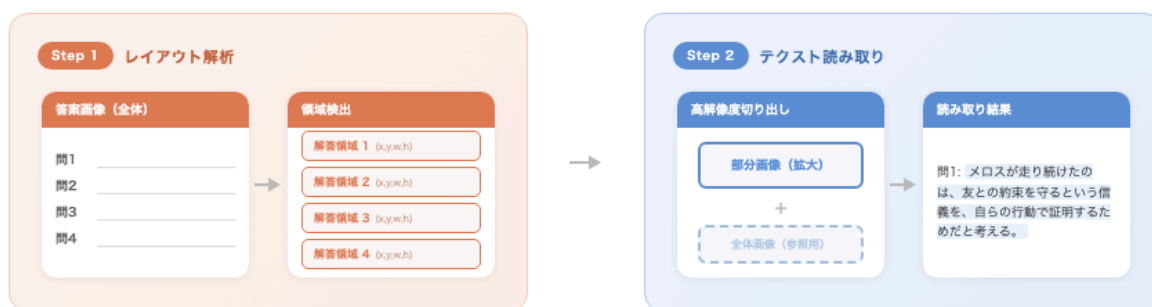


図 3: 空間プロンプティングの処理フロー — 全体画像からの領域検出と高解像度部分切り出しによるテキスト読み取り

OCRが完了すると、本採点の前にループリック精緻化というフェーズが入る。AIが実際の生徒の解答パターンを分析し、「この言い換えは正答として認めますか」「この独自の解釈にはどう対応しますか」といった確認を選択式で教員に提示する。教員の回答は採点基準に補足として反映され、それを踏まえて本採点が進む。実はこれは、多くの学校で教員が試験採点時に実際にやっていることでもある。最初の数枚を見ながら「この書き方は正答に含めよう」「ここは部分点にしよう」と基準を調整していく、あの作業をシステムの中に組み込んだ形だ。

▼ 採点基準を答案から精緻化する (推奨)

AIが実際の学生の解答を読み、判断が分かれそうなケースを具体的に指摘します。事前に回答しておくことで、採点のプレが減ります。

答案を分析して確認ポイントを抽出

✓ 分析完了 (6件の確認ポイント)

全学生の解答を通読し、ボーダーラインケースを抽出しています。

6件の確認ポイントが見つかりました。

問2-3. 独自の解釈

[S002] 「聴くということは相手の言葉をそのまま受け入れることではなく、相手の意見によって自分自身の考え方や物の見方が変化するような体験であるということ。」

要素A (「了解の枠内に回収する」の否定) について、S002は「そのまま受け入れることではなく」と表現しています。これは本文の「既有的理解に同化させる」ことの否定とはニュアンスが異なりますが、独自の解釈として部分点を与えますか？

回答を選択

要素Aの言い換えとして認め、満点 (4点) を与える

不十分な解釈として部分点 (2点程度) を与える

誤読とみなし、要素Aは0点とする

図 4: 採点基準精緻化の画面 — AI が生徒の実際の解答パターンを分析し、教員に選択式で確認を求める

採点ロジックにおいては、キーワードの有無だけでなく概念レベルでの合致を判定する方式を採用している。これは国語の記述式採点としてはむしろ自然な設計だ。さらに、AIが一度採点した結果を別の指示文で再検証させる二重チェック機構や、200枚の答案を通じて採点基準がブレていないかを定量的に監視するドリフト検出も組み込んでいる。いずれも、教員が「AIの判断が割れた箇所」「基準がブレた箇所」だけを重点的に確認すればよいようにするためのものだ。

3.4. 残された課題

ただし、手書き文字の認識精度はまだ十分ではない。高解像度化と領域切り出しによって「閨土」を「関谷」と誤読するような致命的なミスは減ったが、消しゴムの跡、走り書きの崩し字、答案用紙の劣化によるノイズは、現行のモデルでも安定して処理できるとは言いがたい。

加えて、処理速度のボトルネックがある。現在のシステムでは、高精度な推論モデルを使用しているために1人あたりの処理に60～120秒を要し、200人分の答案をOCRにかけると約7時間に達する。高速モデルへの切り替えや並列処理の導入を検討しているが、速度と精度のトレードオフがある。こうした実務上の課題を一つずつ潰していく作業が、現在の開発の中心である。

3.5. 現場への導入

専任の先生方からの反応は上々で、講習会の依頼も受けている。ただし、現実的に言えば、手書き答案の自動採点を定期試験のような成績に直結する場面でいきなり使うのはまだ難しい。まず着手しようとしているのは、Google Classroomなどで回収した授業内課題の採点だ。デジタルテキストであれば手書き認識の不確実性を回避できるし、成績評価の重みも定期試験より軽い。AIによる自動採点に加えて、個々の生徒に対する詳細なフィードバック、つまりより良い記述をするにはどうすればよいか具体的なコメントを自動生成して返却する。こうした運用から始めて、精度と信頼性の実績を積み上げていく計画だ。2026年4月から実験的に運用を開始する。

4. 学術知を動画にする試み

4.1. 学術論文から教室へ

採点支援AIとは別に、もうひとつ並行して進めてきたプロジェクトがある。学術論文の知見を対話形式の解説動画に変換し、反転授業の教材として活用する試みだ。

最初に手がけた「走れメロス」の解説を例にとると、工程は次のようになる。まず安藤宏、佐野幹、幸田国広ら6本の学術論文をOCRでテキスト化し、メタデータとともに整理する。それらを精読した上で、2人のキャラクター（文学に詳しい解説役と、素朴な疑問を投げかける聞き役）による20分の対話台本を構成する。この台本は単なるセリフの羅列ではない。キャラクターの表情

の切り替え、BGMの変更ポイント、画面レイアウトの転換、解説用スライドの挿入指示まで、動画の演出情報をすべて含んだ「完全台本」である。

台本の初稿が完成すると、5つのペルソナによるエージェントチームレビューにかける。カジュアルな視聴者、教養系コンテンツのファン、YouTubeプロデューサー、文学研究者、映像演出家。それぞれの視点から台本を検討させ、改善案を出させる。ひとりの人間が見落とす「伝わらなさ」を、異なる立場のペルソナが拾い上げてくれる。レビューを反映した第二稿を経て、ファクトチェック（引用の原文照合、人名・年号の確認）と整合性検証（用語のブレ、論理の飛躍の検出）を行い、最終的に自分で内容を確認して台本を確定させる。

台本が確定すれば、そこから先は自動化されている。VOICEVOX（無料の音声合成エンジン）がセリフを読み上げ、画像はAI生成とライセンスフリーの素材検索を組み合わせで調達し、必要に応じて図式的な解説スライドも新規作成させる。それらをPythonスクリプトとffmpeg（動画編集ソフト）がタイムラインに沿って自動合成し、BGMの切り替えやキャラクターの立ち絵の表情変更まで、台本の指定通りに組み上げる。



図 5: 動画制作パイプラインのワークフロー — 学術論文から完成動画まで7フェーズの品質管理プロセス

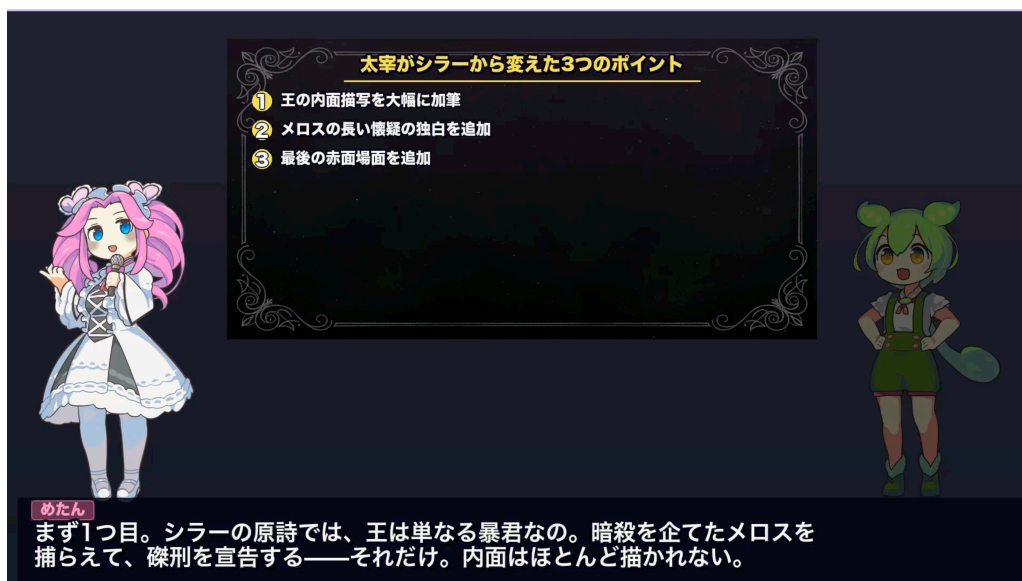


図 6: 完成した解説動画の一場面 — 解説役と聞き役の対話形式による「走れメロス」解説

4.2. 翻訳としての台本構成

このパイプラインの中で最も本質的な作業は、技術的な実装ではなく、「学術知をどう対話に変換するか」という翻訳の問題だ。論文の論旨を正確に把握した上で、視聴者にとって意味のある論点を選び出し、専門用語を平易な言葉に置き換えつつ、学術的な正確性を損なわないラインを見極める。

2人のキャラクターの役割分担もこの翻訳の一部だ。知識は必ず解説役から提示され、聞き役はリアクションや素朴な疑問を返す。聞き役が専門的な知見を先回りして語ると、視聴者にとっての「わからなさ」が消えてしまう。鋭い気づきは、「身近な体験に引きつけた素朴な疑問」の形で表現する。

4.3. 授業への組み込みと専門知の伝達

この動画の第一の用途は、反転授業の教材である。生徒が事前に動画を視聴し、作品の学術的な面白さや新しい読み方を掴んだ上で教室に来る。教室では、動画では扱えない比喩表現の微視的な読み取りや、テキストに即した議論に時間を使える。ただし、動画はあくまで入り口であり、授業のすべてを代替するものではない。

パイプラインが完成した現在、新しい動画の制作に必要なのは学術論文の選定と台本の構成、つまりインプットと編集判断だけになっている。複数の動画企画を並行して進めており、技術的な制作コストは大幅に下がった。

このパイプラインは文学に限定されない。学術論文を選び、論旨を把握し、対話台本に落とし込むという工程は、どの分野の研究者にも応用できる。音声合成や画像生成、動画編集のツールはすべて汎用的なものであり、分野固有の知識は台本構成の段階にしか現れない。つまり、各分野の専

門家が自分の専門知を動画として発信するためのハードルは、すでに大幅に下がっている。学際的な交流の最大の障壁は、専門外の人に内容の質を落とさずに理解してもらうことの難しさにあるが、こうしたパイプラインの普及はその障壁を確実に低くする。

こうした文脈の中で、動画という媒体を選んでいること自体にも意図がある。インターネットにおけるコンテンツ消費の重心が、テキストから動画へと不可逆的に移行しつつあることは、プラットフォームの歴史を振り返れば明らかだ。ブログから SNS へ、SNS から短尺動画へ。情報の流通経路が変わるたびに、そこに乗れなかった知は届かなくなる。学術知も例外ではない。論文や書籍という形式でしかアクセスできない知は、この変化の中で到達範囲を狭めていくおそれがある。動画パイプラインの構築は、学術知を現在の情報流通に適応させる試みでもある。

5. 教育現場の現在地

5.1. ギャップの構造

教育現場における AI 活用の最大の障壁は、技術的な制約ではない。教員が AI で何ができるかを知らないことでもない。正確に言えば、知る余裕がないのだ。

国語科の教員の日常は、授業準備、教材研究、採点、生徒対応、保護者対応、校務分掌と、そもそも新しい技術を試す時間的・精神的な余白がほとんどない。ChatGPT や Gemini のチャット画面を一度は触ってみたという教員は少なくないが、それを自分の業務に本格的に組み込むところまでたどり着く人は稀である。自分の教材を AI に読み込ませて推論させることが可能だということ、Claude Code や Codex のような AI コーディングツールが教員の業務設計にも使えるということ、さらにはそれらが教育のあり方そのものを変えうるということ。こうした認識の層が、少なくとも自分の周囲ではほとんど共有されていない。

これは日本に限った話ではない。RAND Corporation が 2025 年に発表した調査によれば、アメリカでも教員のうち AI を授業に活用しているのは全体の約 25% にとどまる。しかも、貧困率の高い学区ほど導入率が低く、AI 活用に関する研修や指針の整備にも差がある。世界的に見ても、教育現場の AI 活用は「使える人だけが使っている」段階にある。

Khan Academy が開発した AI チューター「Khanmigo」は 380 以上の学区に導入され、成功事例としてしばしば引き合いに出される。その展開において一貫して強調されているのが「教師主導の設計」という原則だ。AI が教師を代替するのではなく、教師が AI の使い方を設計する。この原則には同意するが、問題はその「設計」ができる教員がどれだけいるか、ということだ。

5.2. アプリケーションとして届ける

だからこそ、コーディング AI を使いこなせる少数の人間が、そうでない多数の人間のためにワークフローを設計し、それを GUI つきのアプリケーションとして届けることが大事なのだと思

う。Khanmigoの「教師主導の設計」は正しい原則だが、自分が目指しているのはもっと手前のところで、教員が設計を意識しなくても使える形にまで落とし込んだツールを届けることだ。

採点支援AIのデモを専任の先生方に見せたとき、「こんなことができるのか」という反応が返ってきた。しかし、デモを見せただけでは何も変わらない。先生方が自分の答案を自分でアップロードし、自分の採点基準で結果を得られる。その状態まで持っていかなければ、「すごいね」で終わる。

第3節で述べた採点支援AIは、現在、複数の教員が同時に利用できる環境の構築に着手している。生徒の答案データを扱う以上、技術的なセキュリティだけでなくデータガバナンスの設計が不可欠だ。答案画像を外部のAI（GeminiやClaude）に送信してOCR・採点を行う仕組みである以上、送信前に氏名欄をマスクする処理、保存時の暗号化、保持期間を超えたデータの自動削除といった仕組みを設計に組み込んでいる。プライバシーポリシーの策定も進めているが、法務面の整備はまだ途上であり、学校側と協議しながら4月の実験的運用までに詰めなければならない課題のひとつだ。

一方で、AIへの依存がもたらすリスクにも目を向ける必要がある。Bastaniらが2025年にPNASに発表した研究では、トルコの高校生約1,000人を対象とした実験で、ガードレールのないAIチューターを使った生徒は練習時の成績こそ向上するものの、AIなしの後続テストでは成績が低下するという結果が報告されている。一方、適切にガードレールを設計したAIチューターでは、この学習障害は回避された。つまり、AIが学習を助けるか損なうかは、ツールの設計次第だということだ。この知見は、次節で述べるプラットフォーム構想でも常に念頭に置いている。

6. これからの展望

6.1. 統合プラットフォームとその先

ここまで述べてきた採点支援AIと解説動画パイプラインは、いずれ一つの統合プラットフォームに発展させたいと考えている。採点支援、教材に基づいて回答するチューターAI、授業設計の支援、解説動画の生成。これらが共通の教材基盤を共有し、教員がGUIから操作できる形を目指している。ただし、前節で触れたように、ガードレールの設計、たとえばAIの補助を段階的に減らす仕組みや、AIなしでの振り返りを組み込むことは、プラットフォームの根幹に据える必要がある。

まずは2026年4月から、自分が教えている国語科で採点支援AIの実戦投入を始め、先生方のフィードバックを得る。その後の開発で記述式だけでなく選択式や短答式の問題にも対応させたので、次の段階として他教科の教員にも使ってもらい、教科ごとのニーズを吸い上げていきたい。

ただし、手書きのペーパーテスト全般への展開には根本的な壁がある。漢字の書き取り問題や、短答式であっても解答の漢字が正しいかどうかの判定は、OCR（文字の読み取り）ではなく、書かれた画像そのものの正確さを評価する問題であり、現行の基盤モデルではまだ安定して処理できない。Geminiのようなマルチモーダルモデルの画像認識能力が向上するのを待つしかない部分もあり、ここが手書き採点の展開を左右するボトルネックになっている。

それでも、デジタル入力 of 課題から始めて段階的に対象を広げていくことはできる。AIに応答させるにはコストがかかる以上、持続可能な形で届けるには事業化が必要だ。意思決定の速い私立校を皮切りに、教科を問わず全国の教員が使えるプラットフォームに育てていくことを目指している。

一方、第2節で述べた文学研究のデータベースは、別の形で公開を目指す。著作権法上の整理が必要なため、まずは大学・研究機関向けの提供から始め、法的な検討を経た上で、将来的にはWebサービスとして広く研究者に開放したい。

ただ、この8ヶ月で考えたことの中で本当に引っかかっているのは、プラットフォームの設計よりもっと手前の話だ。人文学の知そのもののあり方が変わりつつある、という感覚がある。

6.2. 閉じた回路を開く

人文学の世界では、ある作品やテキストの「解釈」は、長い時間をかけて文献を読み込み、概念を結びつけ、論として発表するという営みを通じて生産されてきた。そしてその解釈を生産できるのは、基本的に専門の研究者に限られていた。論文を読む人間も、ほとんどが同じ分野の研究者である。率直に言って、その論文が専門分野の外の誰かに届いているかどうかは、いつも心許ない。人文学の知は、意図的に囲い込まれているわけではないが、結果として閉じた回路の中にとどまっている。

この回路が開かれつつある。第2節で述べたように、構造化されたメタデータとテキストデータがあれば、AI エージェントは自律的に文献を横断し、関連する議論を拾い上げ、問いに応じて再構成できる。研究者が何年もかけて培ってきた「この文献とこの文献がつながっている」という知識のネットワークを、不完全ではあるが、エージェントはデータの構造から再構築できるようになりつつある。そうなれば、専門的な知の探索を研究者以外の人間が行うことが、少なくとも原理的には可能になる。

6.3. 道を均す、ただし均し方は中立ではない

ではその変化の中で、専門家には何が残るのか。私は、「目的地を示す案内人」から「人がどこへでも行けるよう道を均しておく人」への移行が起きつつあると考えている。自分がやっていることも、結局は道を均す仕事だと思う。現代詩や近代文学の資料を電子化し、メタデータを付与し、エージェントが検索できる形に整備する。特定の解釈を生産するためではなく、誰かがいつか必要としたときに、その知にたどり着ける道を作っておくためにやっている。採点支援 AI も動画パイプラインも、根底にあるのは同じ発想だ。

ただし、道を均す行為は、中立的なインフラ整備ではありえない。

メタデータのスキーマを設計するとき、どのキーワードを付与するか、どのカテゴリで分類するか、どの概念を「対象作品」としてタグづけするか。そこで私は不可避免的に、日本文学研究が歴史的に作り上げてきた枠組みの上で作業している。論文それ自体が従来の論点や概念の延長線上に書かれている以上、それに付与するメタデータもその枠組みを引き継ぐ。均された道は、特定の学

問的伝統の地図の上に引かれている。社会学者や認知科学者が戦後詩にまったく別の問いからアクセスしたいとき、この道は必ずしも彼らの歩きたい方向を向いていない。

そしてこの限界は、技術的にも具体的だ。メタデータに含まれないキーワードや概念は、カタログのテキスト検索では拾えない。現在の設計では、エージェントがまずカタログを検索して候補を絞り込み、該当する論文の全文を読んで推論する。200巻規模であればこのアプローチで十分に機能するし、将来的にはベクトル検索を併用することで数千から数万冊規模への拡張も視野に入れている。しかし、入口の絞り込みにメタデータを使う以上、スキーマが想定していない切り口の問いには構造的に弱い。全文を読む段階に到達できれば枠組みを越えた発見は起こりうるが、そもそもその論文にたどり着けるかどうか、メタデータの設計に依存してしまう。

つまり、枠組みの外からの問いが立ったとき、それに応えるためにはスキーマ自体を書き換える必要が出てくる。どのフィールドを足すべきか、どの分類軸が必要か。その判断には、問いの文脈を理解し、既存のスキーマの何が足りないかを見抜く知識が要る。「道を均す」仕事の核は、一度均したら終わりではなく、新しい問いに応じて均し方を設計し直し続けることにあるのだと思う。

もっとも、この「設計できる少数の人間」の特権性も盤石ではない。第2節で見たように、AIコーディングツールの進化によって構築の敷居は急速に下がっている。スキーマ設計やパイプライン構築が、いずれ広く共有される技能になる可能性は否定できない。

なお、以上は文学研究と教育の現場から見た話であり、哲学や歴史学、あるいは社会科学の領域では、知の構造化とアクセスの問題がまた別の形をとるだろう。

6.4. 問いを持ち込む人、テキストに戻る人

それでも、道が均されることの意味は大きい。エージェントは問いそのものを生み出すわけではないが、人間の側に問いさえあれば、それを既存の枠組みの外に連れ出す力を持っている。文学研究の訓練を受けていない人が「走れメロス」に対してまったく別の角度から問いを持ち込んだとき、エージェントはメタデータの網目をくぐりながら、予想外の文献のつながりを見つけ出さう。問いの力によって枠組みが更新されていく。それは、良質な研究と呼ばれてきた営みと地続きのことだ。

ただし、AIはもっともらしい補助線を引く力を持つがゆえに、根拠の薄い推論も同じように補強できてしまう。生産的な読みと暴走を分けるものは何か。私は、テキストに立ち戻れるかどうかだと考えている。

授業をしていると、そういう場面に何度も出くわす。生徒が飛躍した解釈を書いてきたとき、結局やることは「でも本文にはこう書いてあるよね」とテキストに引き戻すことだ。テキストという動かない対象に突き合わせることで、読みは根拠を問われ、修正され、あるいはかえって強化される。そして、テキストが検索可能な形で誰でもアクセスできる状態にあること、つまり資料のデジタル化とメタデータ付与の仕事が、この検証の回路を広く開くことにほかならない。道を均すという営みは、知へのアクセスと知の検証可能性を同時に担保する。

5年間、中高生に国語を教えてきて実感していることがある。生徒に作品の解釈を書かせると、9割はテキスト本体から簡単に反証できてしまうような飛躍した読みだ。しかし、たまに、本当にた

まにだが、こちらが考えもしなかった鋭い問いの芽が混じっている。その芽は、その生徒自身の経験や感覚から生まれたものだ。問題は、根拠の示し方も関連する議論の在処も知らないために、鋭い直感が「感想文」のまま消えていくことにある。AIが関連する学術的議論を引いてきて「その方向には、実はこういう議論がある」と補助線を示せば、初学者の直感は検証可能な読みへと育つ可能性がある。ただし、補助線を引きすぎれば生徒自身が問いを手探りで育てる過程を奪いかねない。その設計判断は慎重に行っていくしかない。

とはいえ文学は、正しさがすべてではない。ある人間がある作品を読んで、たとえそれが学術的には誤読だったとしても、その読みがその人の何かを変えたなら、それだけで文学の価値はあったのだと思う。道が均され、ツールが敷居を下げ、テキストへのアクセスが開かれた先で、一人ひとりの問いと読みが立ち上がる。AIの時代における人文学は、閉じた回路の中で知を精製し続けることではなく、知への道筋と検証の手段を整え、そこから先の探索を多くの人に開いていくという形で、社会の中に存在し続けるのではないか。

7. おわりに

本稿では、基盤モデルの急速な進化に振り回されながら、近現代詩のデジタルアーカイブ構築、採点支援 AI の開発、学術知の動画化、そして新たな OCR パイプラインの構築に取り組んだ 8 ヶ月間の実践を報告した。

数ヶ月かけて作ったものの最適なやり方が翌月には変わり、2日で作ったものが数ヶ月の改良を要し、先月の最善手が今月にはもう古い。足場が動き続けるなら、足場に依存しない立ち方を覚えるしかない。手法は変わる。ツールも変わる。しかし、「この資料はこう整備すれば使いやすくなる」「この採点基準は生徒の実態に合っていない」「この論文の論旨はこう要約すれば伝わる」といった、現場を知っているからこそ出てくる判断は、そう簡単には陳腐化しない。

そして、道を均すことは思っていたより複雑な仕事だった。均した道は中立ではなく、既存の学問的枠組みの方向を向いている。それでも、問いを持った人間がその道の上を歩き出すとき、道の設計者が想定しなかった場所にたどり着くことがある。その可能性を閉ざさないこと、知への道を開き、テキストへの検証可能性を担保し、そこから先の探索を人に委ねることが、いま自分にできることだと思っている。

4月から先生方の手元に採点支援 AI を届ける。そこから何が起きるかは、まだわからない。

参考文献

- Bastani, H. et al. (2025). Generative AI without Guardrails Can Harm Learning. Proceedings of the National Academy of Sciences. <https://doi.org/10.1073/pnas.2422633122>
- Li, X. et al. (2025). Long Context vs. RAG for LLMs: An Evaluation and Revisits. arXiv:2501.01880.

- RAND Corporation (2025). Uneven Adoption of AI Tools Among U.S. Teachers and Principals. ERIC: ED672957.
- Singh, A. et al. (2025). Agentic Retrieval-Augmented Generation: A Survey on Agentic RAG. arXiv:2501.09136.