



配布先: 文部科学記者会、科学記者会、名古屋教育記者会

2024年5月30日

報道機関 各位

生成 AI でソクラテスに人生相談できる！？ 西洋古典学の飛躍的發展に寄与する対話システムを開発

【本研究のポイント】

- ・生成 AI ^{注1)}技術を用いて西洋古典学の研究と教育に新たなアプローチを提供する「ヒューマニテキスト」(Humanitext Antiqua)を開発。
- ・西洋古典原典のデータベースを再整備し、生成内容の出典情報をユーザーが常に参照できる信頼性の高い専門的解釈システムを構築。こうした試みは国内外でも初めて。
- ・「ヒューマニテキスト」は言葉の意味を基にした文脈探索が可能。専門外の人々も興味に合わせて文学・歴史・哲学の原典を気軽に参照できるように。

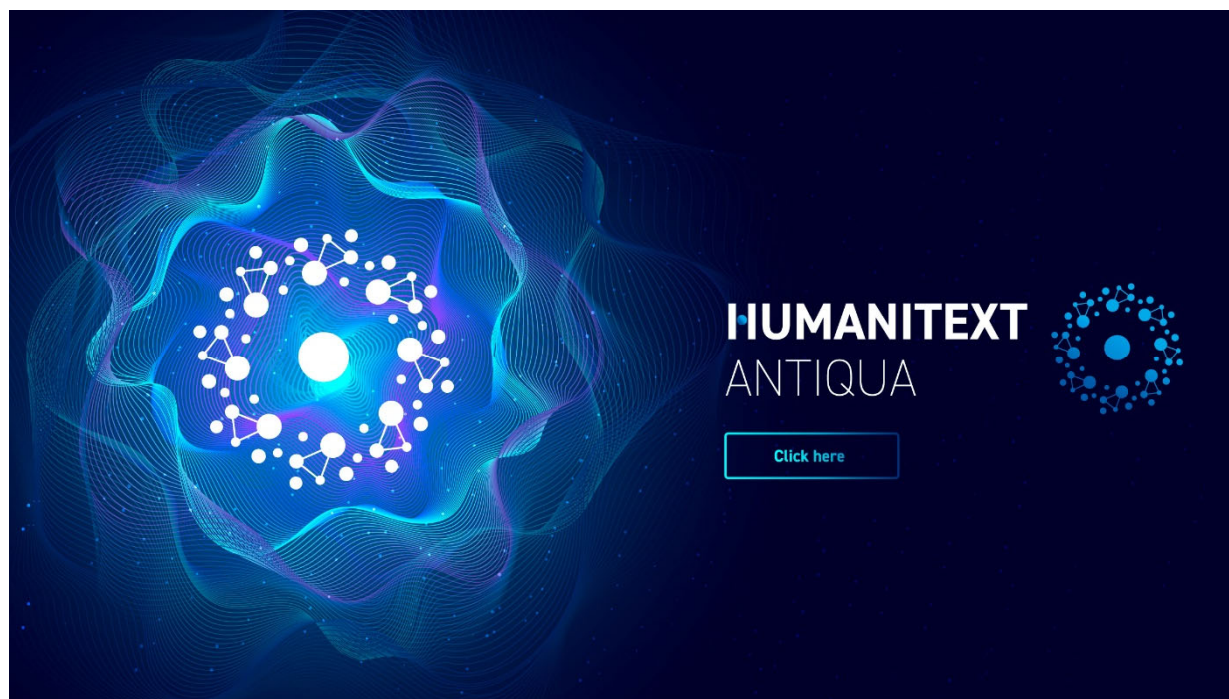
【研究概要】

名古屋大学デジタル人文社会科学研究推進センターの岩田 直也 准教授、桜美林大学の田中 一孝 准教授、国立情報学研究所／ROIS-DS 人文学オープンデータ共同利用センターの小川 潤 特任研究員らの研究チームは、生成 AI 技術を用いて西洋古典学の研究と教育に新たなアプローチを提供する「ヒューマニテキスト」(Humanitext Antiqua)を開発しました。

この研究の重要な成果は、生成 AI が直面する偽情報生成問題に対処したことです。研究チームは西洋古典原典のデータベースを再整備し、学術的水準を保った回答を出力するシステムを開発しました。さらに、生成内容の出典を確認できる機能も搭載しています。このような試みは国内外でも初めてのものです。

「ヒューマニテキスト」は、言葉の意味に基づく文脈での探索を可能にすることで、専門外の人々でも効率的に文献を見つけられます。将来的には、データ基盤のさらなる拡充を図り、時代や文化、学問分野を超えた新しい知識の発掘を目指していきます。

「ヒューマニテキスト」は 2024 年 6 月 1 日から 2 日の日本西洋古典学会で限定公開され、その後一般公開される予定です。今後このプラットフォームには、論文データや辞書、翻訳、註釈が組みこまれ、人文学の研究に最適化されるよう進化し続けます。また、教育機関での学習ツールや公共施設での情報提供手段、創作活動へのインスピレーション源としても期待され、西洋古典の普及に重要な役割を果たす見込みです。



【研究背景】

2022年11月に ChatGPT が登場して以降、文章やプログラムコードなどを生成する AI 技術が世界中で注目を浴びています。その核にある大規模言語モデル(LLM)^{注2)}とは、学習したデータに基づき言語の出現確率を予測する人工知能です。

OpenAI 社が 2023 年 3 月に公開した LLM である GPT-4 は、古代ギリシア語やラテン語を含む多言語での文脈処理と文章生成において非常に高い精度を実現し、人文学分野での活用可能性を急速に高めました。これまで人間にのみ可能だと思われてきた人文学のテキストを機械が「読解、解釈、分析」という革命的な事態が急に現実化したのです。LLM や生成 AI の技術を人文学の研究と教育にどのように生かしていくかという問題は、私たちが真剣に検討すべき喫緊の課題となっています。

人文学分野への LLM や生成 AI の導入は、他の学術分野に比べて著しく遅れをとっています。その背景には、LLM が使用する学習データの信頼性への懸念や、誤った情報を生成するリスク、対話型 AI に適切な指示を与えるスキルの必要性といった課題があるからです。しかし、信頼できるデータ基盤に基づいた研究手法や倫理規定、教育システムがしっかり整備されれば、人文学分野でも AI 技術の導入は十分に可能です。そこで私たち研究チームは、これらの課題を解決し、人文学分野で LLM の多言語による文脈処理の能力を駆使した新たな研究方法を提示することを試みました。

【研究内容】

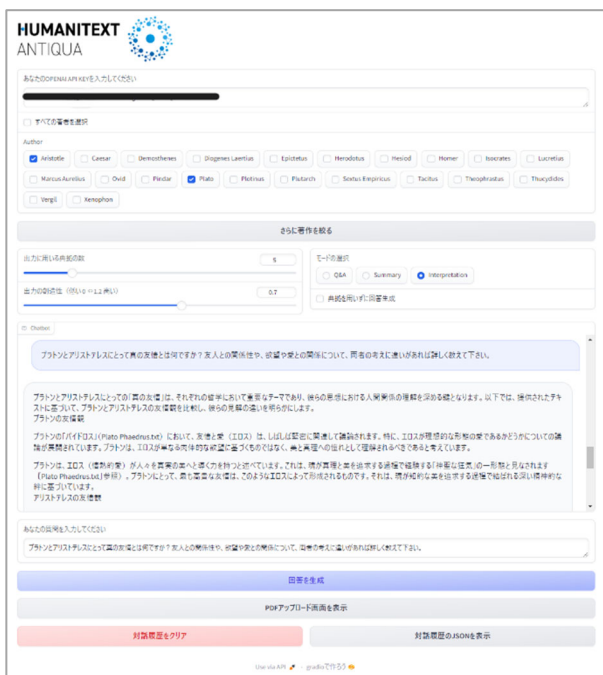


図1:ヒューマニテキスト操作画面

この研究背景を踏まえ、私たち研究チームは、西洋古典学に特化した AI 対話システム「ヒューマニテキスト」(英名:Humanitext Antiqua)を開発しました(図1参照)。

ヒューマニテキストの特徴は、西洋古典分野で広く利用されている Perseus Digital Library^{注3)}におけるオープンソースのテキストデータなどを再構成することで、信頼できる原典テキストに基づいた回答を生成できることです。さらに、回答の典拠となる原典テキストとその典拠情報も同時に出力することができ、ユーザーは出力の正確性をいつでも容易に確認することができます。

このような仕方で、偽情報生成(ハルシネーション)の問題を可能な限り低減させること

で、ユーザーが安心してヒューマニテキストを使用することができます。

ヒューマニテキストの活用によって、これまで個々の著者や著作の研究に細分化されてきた西洋古典分野において、領域横断的な研究が加速度的に進展することが期待できます。

現在、古代ギリシア語やラテン語の古典テキストを研究する際に使用されるデータベースは、特定の単語の用例を検索することはできるものの、そのデータ量が膨大であるため、すべての検索結果を一つ一つ調べるのはほぼ不可能です。このため、研究者は特定の著者や著作に調査対象を絞り、手作業で用例を確認するしかありませんでした。

しかし、ヒューマニテキストは、テキストの文脈を高度に理解できるため、ユーザーが自らの興味のある事柄について母語で質問するだけで、それに関連する原典テキストを特定した上で回答を出力できます。

これにより、古代ギリシア・ローマの哲学、文学、歴史文献をより広い視点で比較しながら研究することが可能となります。この新しいテキスト探索・分析手法は、西洋古典学だけでなく、人文学全体で幅広い比較研究を進めるための大きな一歩となるでしょう。

Press Release

以下にヒューマニテキストの機能を簡単に紹介します。

① 著者および著作選択機能

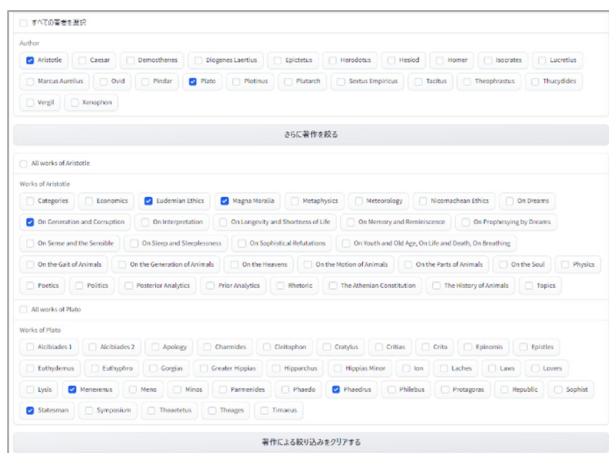


図2: 著者および著作選択画面

プラトンやアリストテレスの哲学的著作を含む 22 人の著者の全作品が利用可能です。一般公開に向けて、西洋古典の代表的な著者すべての作品との連携を完了する予定です。

② 出力カスタマイズ機能



図3: 出力カスタマイズ画面

次に、ヒューマニテキストからの出力を様々にコントロールできる設定について説明します(図3参照)。

まず「出力に用いる典拠の数」では、ユーザーの質問に対してどのくらいの文脈数を探索して回答に用いるか設定できます(現在 1 から 25 の範囲)。また、「出力の創造性」では、設定を低くすると探索した文脈に忠実な回答が出力され、高くするとその文脈に基づきつつもより創造的な回答が得られます。

さらに、「モードの選択」によって、ユーザーの多様な要求に対応できます。「Q&A」モードは、特定の事柄についての情報とその出典を知りたいときに便利です。例えば、ディズニー映画の題材にもなっているギリシア神話のヘラクレスがどのような怪物と戦ったのか、または漫画『テルマエ・ロマエ』を読んで古代の人々がどのようにお風呂に入っていたのかを調べたいとき、このモードを使えば出典付きの回答が得られます。

「学習」モードでは、調べたいトピックに関して、様々な原典の内容を比較しながら、より包括的で詳細な回答が得られます。例えば、「カエサルはなぜ暗殺されたのか」という質問に関して、様々な歴史的記述を比較検討できます。このモードでは、参照する典拠をすべてユーザーが使う言語に翻訳し、予備知識も補足して回答が出力されるため、学習者に最適です。

「専門解釈」モードでは、質問に関連する原典テキストの言語的分析や深い解釈にまで踏み込んだ内容を出力できます。例えば、「プラトンやアリストテレスは『知』をどのように捉えていたか」といった哲学的な質問に対して、古代ギリシア語の原典を参照しながら専門的で高度な回答が得られます。

「対話」モードは、ヒューマニテキストの魅力の一つで、一般の人も楽しめるモードです。このモードでは、選択した著者の原典に基づいて、その著者や登場人物と実際に会話しているかのようなやり取りができます。例えば、哲学の祖とされるソクラテスに「幸せになるにはどうしたらよいか」と質問したり、『自省録』で有名なストア派のマルクス・アウレリウスからストイックに生きるための指南を得ることも可能です。

最後に、「典拠を用いずに回答生成」欄にチェックを入れると、ChatGPT のように、LLM の学習データを利用しての回答に切り替えることができます。この選択肢の活用は多岐に渡り、事前に得た回答を西洋古典分野外の広い文脈から捉え直すことや、古代ギリシア語やラテン語を日本語に翻訳してもらうことなど、ユーザーのニーズに応じた様々なタスクを遂行することが可能です。

③ 典拠確認機能

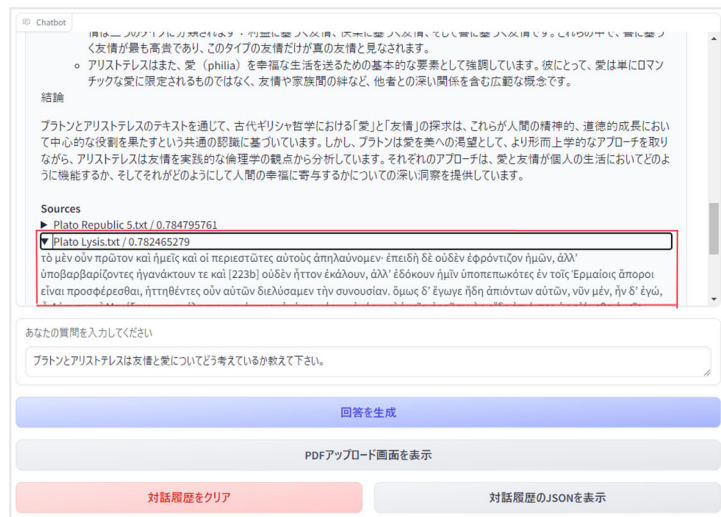


図4:チャット画面

最後にチャット画面の紹介です(図4参照)。

まず、ユーザーが質問すると、その回答が順次生成されます。そしてその最後に、赤枠で囲った欄のように、その回答に用いられた典拠が Sources として一覧で表示され、クリックすると原典テキストを確認できるようになっています。この原典にはテキスト典拠情報が付与されているので、ユーザーは簡単にその箇所を特定することができます。

また、ヒューマニテキストは、単なる Q&A システムではなく、チャット機能を備えているので、過去の対話履歴を踏まえた上で、新たな文脈探索および回答の出力を行えます。

さらに、右下の「対話履歴の JSON を表示」ボタンをクリックすれば、その対話履歴データを典拠情報も含めた上で、外部アプリケーションでの再利用等に適した JSON 形式で表示することができます。

【今後の展開】

このように「ヒューマニテキスト」は、西洋古典原典へのアクセスのハードルを大幅に下げることによって、専門家以外の研究者や学生、一般の人々も日本語で西洋古典のリソースを気軽に引用・参照・解釈できる環境を提供します。西洋古典は、西洋や人類全体の知の源泉であり、このシステムを通じて多くの人々がこの過去の遺産からインスピレーションを

得る機会を広げたいと考えています。

また、教育現場では、これまでの文献精読を中心としたトレーニングではなく、創造的に問いを立てる課題発見能力と AI を適切に運用する能力の育成により大きな比重を置く必要が想定されます。本研究チームは、「ヒューマニテキスト」を活用したそのための教育プログラムの開発にも取り組んでいます。

さらに、今回開発した新しいテキスト分析システムは、テキストデータさえ整備されれば、人文学全体で幅広く応用可能です。この成果は、西洋古典学の研究を新たな次元に引き上げるだけでなく、あらゆる時代や地域の哲学・思想、文学・言語、歴史・文化にわたるテキストを包括的に分析し、新たな知識の開拓と理解の深化に貢献します。

これまで専門分野が細分化されてきた結果、各分野の知識は蓄積されてきた一方で、分野間の対話や包括的な視点に立つ研究が難しくなっていました。しかし、「ヒューマニテキスト」を拡張することで、古今東西の知識を統合し、異なる学問分野の理解や思考法が交差する場を作り出すことが可能です。このシステムを活用することで、人文学研究における新たな学問領域や融合領域の開拓が今後の課題となります。

本研究は、2024 年度から始まった萩原学術振興財団第 4 回研究助成の支援のもとで行われたものです。萩原学術振興財団の寛大なご支援に深く感謝申し上げます。

【用語説明】

注 1)生成 AI:

人工知能の一種で、テキスト、画像、音声などを生成する能力を持つ。ChatGPT などがその代表例である。

注 2)大規模言語モデル:

大量のテキストデータを用いて訓練された AI モデルで、人間の言語を高精度で理解し生成することができる。

注 3)Perseus Digital Library:

西洋古典の原典をオープンリソースとして広く公開するプロジェクトで、古代ギリシア・ローマのテキストや関連資料を無料で提供している。

【論文情報】

雑誌名:人工知能学会全国大会論文集

論文タイトル:大規模言語モデルを活用した西洋古典研究と教育

著者:岩田直也¹, 田中一孝², 小川潤³

1 東海国立大学機構 名古屋大学

2 桜美林大学

3 ROIS-DS 人文学オープンデータ共同利用センター

【ヒューマニテキストX(旧 twitter)ページ】



このページでは、ヒューマニテキストの多様な利用例を数多く紹介しています。

URL: https://x.com/humanitext_x

【研究者連絡先】

東海国立大学機構 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 岩田 直也 (いわた なおや)

TEL:052-789-2211

E-mail:n.iwata@nagoya-u.jp

桜美林大学 リベラルアーツ学群

准教授 田中 一孝(たなか いっこう)

TEL:042-797-7985

E-mail:ikkotana@obirin.ac.jp

国立情報学研究所／ROIS-DS 人文学オープンデータ共同利用センター

特任研究員 小川 潤(おがわ じゅん)

TEL:03-4212-2164(国立情報学研究所広報)

E-mail:jun_ogawa@nii.ac.jp

【報道連絡先】

東海国立大学機構 名古屋大学広報課

TEL:052-558-9735 FAX:052-788-6272

E-mail:nu_research@t.mail.nagoya-u.ac.jp

学校法人桜美林学園 総合企画部 広報課

TEL:042-797-9772 FAX:042-797-9829

E-mail:webadmin@obirin.ac.jp

大学共同利用機関法人 情報・システム研究機構 国立情報学研究所

総務部企画課 広報チーム

TEL:03-4212-2164

E-mail:media@nii.ac.jp