

「AIが解説!新潟大学の研究」サイトURL・二次元コード& COAR2025でポスター発表しました

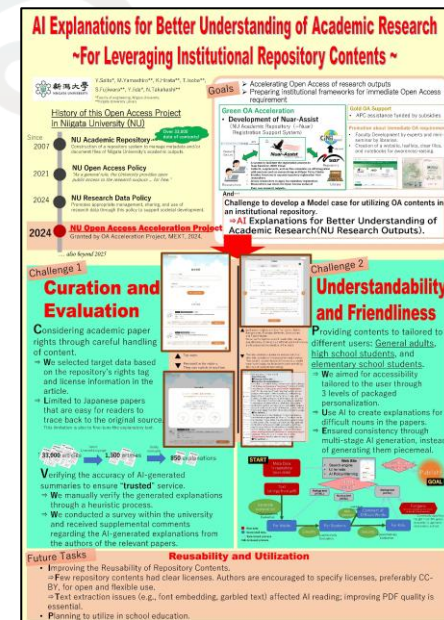
- 「AIが解説!新潟大学の研究」

<https://repolab.lib.niigata-u.ac.jp/>



- COAR2025ポスター発表 (JPCOARリポジトリで公開中!)

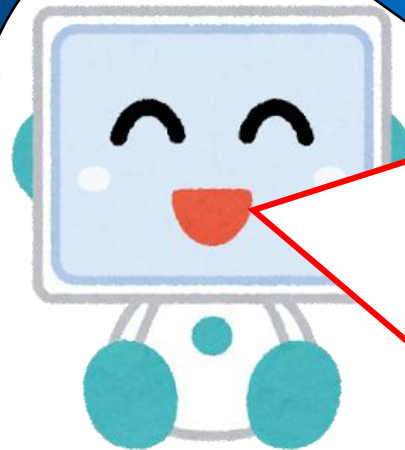
<https://doi.org/10.34477/0002000614>



小学生でもわかる！？AIが“解説”する大学の知 「AIが解説！新潟大学の研究」



生成AIによる
解説文生成



AIが解説！新潟大学の研究



論文のような新潟大学の研究成果は機関リポジトリでオープンアクセスになっているが、内容が難しいものが多く、広く一般の方へ発信することが課題に…

だれでも新潟大学の研究成果に親しめるように、AIが大人向け、中高生向け、小学生向けの3段階でわかりやすく研究を解説！

※新潟大学学術リポジトリに収録されている一部の研究成果を対象に、解説文を生成・Webサイト上で公開しています。

※解説文は都度生成ではなく、事前に生成されたものを表示しています。

2つの壁 ①機械可読性の壁

論文PDFファイルをそのままAIに読ませようとすると、以下のような原因で正常に内容を読めず、正確な解説文を生成できない場合があった。

- フォントがなく、ただの画像として表現されている
 - 一部または全部のフォントが埋め込まれていない
 - 文字をコピーすると化けるor消えるor二重になるor全ての文字間にスペースが入る
 - 文字をコピーすると空白であるはずの別の箇所まで選択される
 - 組版不良など、不適切な段落構成によりコピーに不都合が生じる
 - 行番号がある
 - PDFへの変換時において文字が康熙部首やCJK部首補助に化ける
- …etc.

対応策：読み取り可能なテキストにするため、文字起こしを業者に委託した。

2つの壁 ②著作権（ライセンス）の壁

- 今回は日本語論文を解説文生成の対象にしたもののCC-BYライセンスがついているものが極端に少なかった。
→しかし、構築時に著者全員に逐一許諾を取っている時間が取れず、オプトアウト方式でのみ許諾の確認を行った。
- [著作権法第30条の4](#)や[文化庁「AIと著作権について」](#)ページにある資料等を参照し、以下のように著作権に配慮する形で進めた。

「AIが解説!新潟大学の研究」-このサイトについて より(色、下線は藤原による改変)

(前略)AIは様々な用途で使用されており、機械学習用データとして他人の著作物を利用することは認められています(著作権法第30条の4等)。ただし、当該著作物に表現された思想又は感情を享受する目的での利用は、著者の許諾無しにはできません。この事業では、既存のAIを活用し、モデル入出力レベルでのチューニング(以下、チューニング)によって出力制御を行っています。AIの追加学習は行っていません。さらに、論文の改変ともなりかねない「要約」にならないように意図的なデータ処理及びチューニングを行い、論文の背景・目的等の解説を目的とした「解説文」を提供するようにしています。より具体的には、著作権者の権利を侵害しないことを重視しつつ、原文の不要な引用の回避・省略、専門用語の説明文の埋め込み、ふりがな(ルビ)の付与といった依拠性の回避と独立した創作物となるよう配慮しています。論文の正確な内容については各解説ページのリンクから元の論文を参照してください。

この取り組みから見えた リポジトリ収録論文のAIによる二次利用の課題

- 論文PDFの品質のばらつき

→フォントの埋め込みや文字化けなどのテキスト抽出の問題がAIの読み取りに影響を与えている。

→が、PDFを作成する環境や手順はそれぞれ異なり、著者や印刷所に「品質の高いPDFを作ってください」と徹底することも難しい。

- CCライセンス付き論文の少なさ

→多くのリポジトリ収録論文にはCCライセンスがついていない。

→オープンで柔軟な利用を促進するために、著者は可能なら論文をOAにし、さらにCC-BYライセンスを付与する必要がある。

→ただし、CCライセンス付与の強制はできない。