

図書館と生成AI

ChatGPT等の活用で図書館業務の効率化はなるのか

八洲学園大学生涯学習学部 教授
原田隆史

ushi@e.email.ne.jp

生成AIとChatGPT

人工知能の歴史(ブーム)

- 第一次AIブーム(1960年頃~1970年代半ば)
 - 推論と探索に基づく判断
 - 対話できる自然言語処理プログラム ELIZA
 - 複雑な要因が絡み合う課題を解くことは困難で失速
- 第二次AIブーム(1980年代)
 - エキスパートシステムが事業に広く導入され始める
 - ディープラーニングの基本「誤差逆伝播法」も発表
 - 知識記述の手間, 矛盾するルールに対応できず失速
- 第三次AIブーム(2000年代後半から?)
 - 機械学習、とくにビッグデータ&ディープラーニング

第1次・第2次ブームのAI

- 人間社会の専門家の知識・ルールをコンピュータに教え込む
- エキスパートシステム
 - 専門家の真似をするシステム
 - 大量のYes/Noを教え込むことで解を得る
 - If/Thenの条件分岐を山ほど入れて挙動させる
- ただしうまくいかなかった

従来のAIとエキスパートシステム

- 第一次および第二次AIブーム
 - コンピュータに人間の専門家の知識とルールを教える
 - 専門家の意思決定を模倣するシステム
 - 一連の「はい/いいえ」の質問を通じて解決策を導き出す
 - 多数の「もし～ならば」分岐に基づいて動作する
- 限界と課題：
 - 膨大なルールの中での複雑さと矛盾
 - 広範なルールを管理することが難しい
 - 限られたシナリオでのみ効果的
 - 曖昧さや現実世界の新しい問題への適応に苦勞

ルールベースから統計的機械学習へ

インターネットの普及, コーパス拡大, 計算資源の向上

SVMやベイズモデル(分類・回帰), CRF(系列ラベリング)などを, 人手設計の特徴量と組み合わせ画像・音声・テキストの識別・推定を高精度化する枠組み整備

ルールや典拠に基づく記号的処理を前後処理や制約として用い, 分類・回帰などのスコアリングは統計モデルが担うといった併用が定着

多層ニューラルネットの再評価, GPU計算の普及, 大規模データセットの整備, 深層学習が実用段階へと移行

「Googleの猫」(2012年)

- 特定の概念(例えば「猫」)についての事前知識なしに, 数百万のYouTube画像から学習
- パターンを認識し, 深層学習技術を用いて自発的に分類を行う
- 重要な点は, AIが「猫とは何か」を明示的に教えられず, 自立して特徴を「グループ化」または「認識」した
- AIがグループ化を行った後, 人間はその識別されたグループに「猫」といった用語をラベル付けできた



大森あつし. 2012年にAIの歴史が動いた! ついに猫認識に成功した「Googleの猫」
<https://diamond.jp/articles/-/165402>

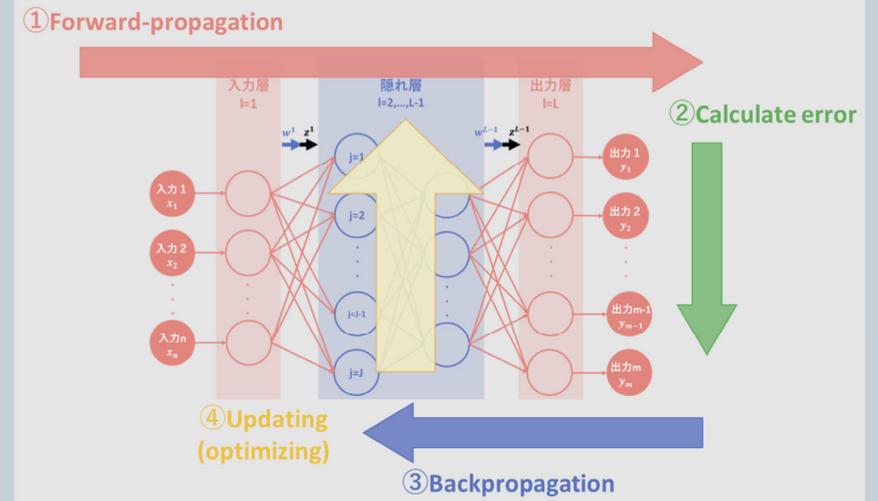
第3次AIブームの背景

- ビッグデータ
 - 各種システムからデータの自動取得が可能になるなど大量のデータが入手できるようになってきた
- 機械学習の実用化
 - 決定木, サポートベクトルマシン(SVM), ランダムフォレスト, 回帰分析など各種の統計的手法をベースに, 入力データが持つ各種特徴量に基づいて出力データを予測する技術が進歩した
- ディープラーニング
 - 脳神経回路の働き(ニューロン)をモデルに, 特徴量が何かわからなくても入力から出力に至る経路を推定できるようになった

非生成AI生成AI

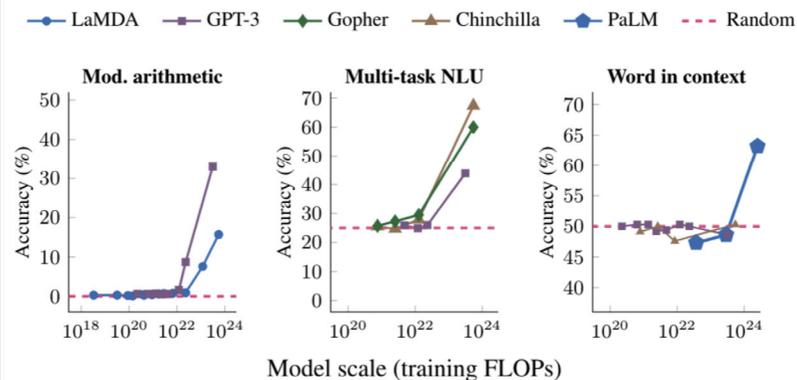
- 学習手法
 - いずれも機械学習, とりわけディープラーニング(深層学習)が土台
 - 非生成AIは, ロジスティック回帰・サポートベクタマシン・勾配ブースティングといった非DL手法や, 規則ベース/知識ベース推論など機械学習に依存しない手法も重要
- 機能
 - 非生成AIは識別・推定
 - 生成AIは表現生成
- 適用対象
 - 非生成AIは自然文生成を伴わずに根拠抽出や選定理由の整形を担う
 - 生成AIは制約付きデコードや検証器の併用により分類・同定の確信度向上に寄与
- 両者の学習目標と入出力の性質は異なることに留意が必要

ディープラーニング(ニューラルネットワーク)



Sakigakeman. 初心者の初心者による初心者のためのニューラルネットワーク
 (https://qita.com/sakigakeman/items/3fc463d8a9613c13c00a)

大規模のデータを使うと急速に賢くなるポイントが



多段階の算術演算を実行する能力(左)、大学レベルの試験に合格する能力(中央)、文脈内の単語の意図された意味を特定する能力(右) Source: "Emergent abilities of large language models." arXiv preprint arXiv:2206.07682 (2022).

ディープラーニングだけが理由ではないが...

- 意味や理由はわからなくても, 何故か似ている箇所を自動的に見つけ出させた
- 超大量・超多数のデータ(ビッグデータ)をコンピュータに投入して, 各種の要素を片っ端から(正確には工夫しながら)組み合わせてルールを発見
- いくら考えても理解できないルールもOK
- 人間社会の専門家の知識・ルールは確認などで使うだけ

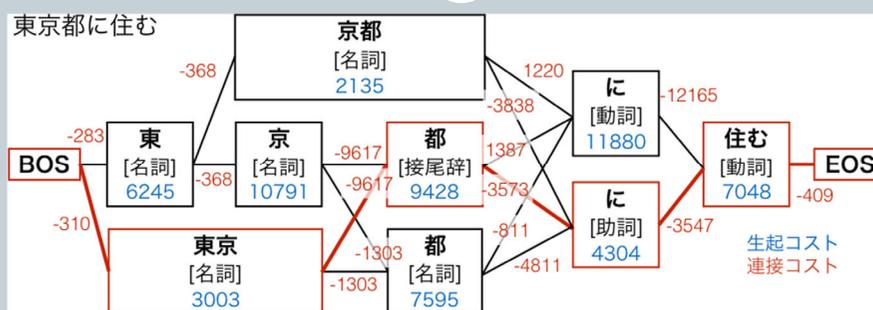
ディープラーニングとAI

- ディープラーニング：脳の神経回路を模したニューラルネットワークにより、中間層を多数重ねて表現を自動獲得する手法
- 入力層-中間層-出力層の結合重み(パラメータ)を誤差逆伝播で最適化し、自動的に高次の特徴を抽出
- 画像認識や自然言語処理、シミュレーションなどで高い表現力と予測力を発揮
- パラメータの可読性は低く、説明や検証は別の仕組み(評価指標、プロンプト設計、根拠提示、ルール制約、人的レビュー)で補う必要がある
- 生成AIのLLMも、非生成AIの高性能な分類器も、この同じ深層学習の土台の上に構築されている点に留意が必要

生成AI(Generative AI)

- ビックデータを元に、機械学習やディープラーニングの手法を駆使して、入力した文字列/画像/音声などに馴染むデジタルコンテンツを生成する技術
- 大量データをもとにしたパターンの組み合わせにより、オリジナルのコンテンツにはないテキストなどを創出することが可能
- 「新しいコンテンツを作り出す能力がある」と表現されることもあるが…

言葉が次につながる例 (日本語変換システム)(AIの説明ではない)



BOS: beginning of sentence
EOS: end of sentence

考えられる解を全て表示したもののデータをラティスという
経路上のリンクのコストとノードのコストの和が最小化されるように
経路探索する、このようなアルゴリズムをビタビアルゴリズムという

<http://techlife.cookpad.com/entry/2016/05/11/170000>

非生成AIと生成AI

- 非生成AI
 - 現実のデータに対して「ラベルを当てる・値を推定する」など識別・推定
 - 評価は適合率・再現率・F1やRMSEなどタスク固有の指標で行い、しきい値や業務ルールと組み合わせで安定運用するのが通例
 - 図書館領域では、典拠コントロール、名寄せ、重複排除、主題候補抽出、所蔵横断の同定、貸出データに基づく需要予測などが期待される
- 生成AI
 - 入力に応じた新しい表現(テキストやコード、画像、音声等)の生成が目的
 - 与えられた文脈に対してもっともらしい次語を逐次選び文章全体を構成
 - 評価は事実整合性・一貫性・有用性・安全性など複合尺度
 - 図書館では、要約・説明・文面作成・問い合わせ応答の草案など表現生成の場面で効果を発揮
- 両者は対立せず、学習基盤(機械学習・深層学習)を共有。目的が異なる

対話AI, 生成AI, 大規模言語モデル

- 対話型AI
 - 対話を模倣してコミュニケーションを進化させるAI
 - 近年は生成AIをベースとするものが多い
- 生成AI
 - 画像や音声、文章などのデータを生成するAI技術の総称
 - 近年は大規模言語モデルをベースとするものが多い
- LLM(大規模言語モデル)
 - 大規模なテキストデータを用いた言語モデルのひとつ
 - 近年はディープラーニング/ニューラルネットワークを用いるものが多い



「私たちの使命は、汎用人工知能 (AGI) が
人類全体に利益をもたらすことを保証することにある」

2015年12月	イーロン・マスク、サム・アルトマンらにより設立
2018年2月	イーロン・マスク、役員辞任
2018年6月	GPT発表
2019年2月	GPT-2発表
2019年7月	マイクロソフトから10億ドルの投資
2020年4月	音楽生成Jukebox発表
2020年5月	GPT-3発表
2022年11月	ChatGPT発表
2023年1月	マイクロソフトが100億ドル投資すると発表
2023年3月	GPT-4発表

ChatGPT (2022年11月)

- Webから収集した45TBのデータセットから570GBのクリーンテキストを選択。このコーパスをもとに教師無し機械学習を適用
- 当初はGPT-3, 現在はGPT-5
- GPT-3は1750億のパラメータを使用しており、GPT-4は1兆規模、GPT-5はそれ以上？
- 一般的に、パラメータが多いほど、パフォーマンスは高い

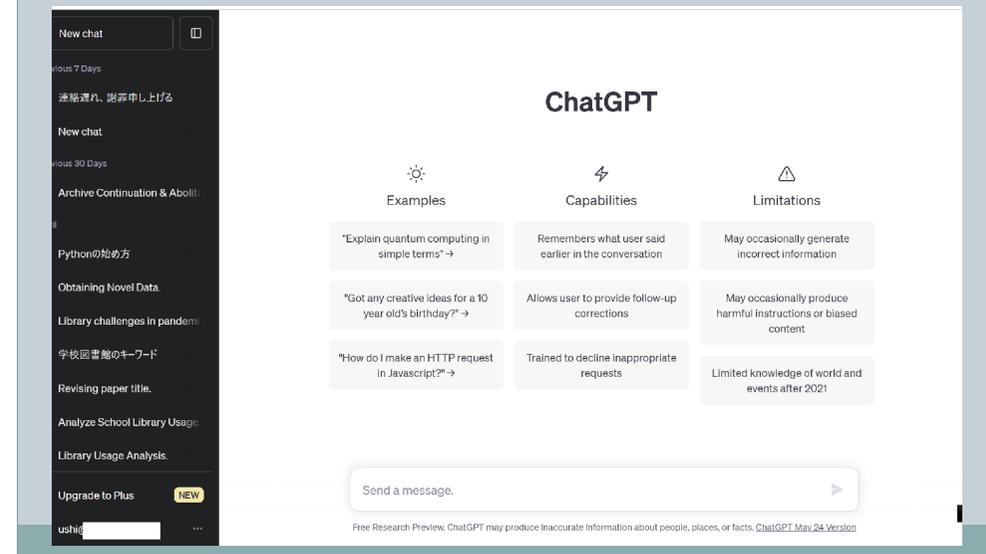
GPT3からInstructGPTへ

- GPT-3/4/4oは人間らしいテキストを生成するが、人間の意図を反映しないことも、時には不正確な内容を作成することもある：AIアライメント
- InstructGPTは、AIの出力を人間の価値観に合わせるために、人間のフィードバックを統合することでこの問題に対処。このプロセスに「人間のフィードバックによる強化学習 (RLHF)」が用いられる
- RLHFは、人間の評価と報酬モデルの両方を用いてAIをファインチューニングし、その応答の質を向上

InstructGPT から ChatGPT

- ChatGPTは基本的にInstructGPTそのもの
- 自然な会話生成のための追加の強化が施された
- 広範な人間とAIの相互作用のサンプルをもとにして、機械学習とファインチューニングによりインターフェースとしての会話テキスト生成能力が向上している
- ChatGPTは主に2021年までに収集された大規模なデータセットをもとに、「人間の好み」に合わせた回答するためにファインチューニングされたもの

ChatGPTで質問する画面



YU 帰無仮説を、中学生向けに説明してください。

帰無仮説とは、統計学で使われる考え方の一つです。何かを調べるとき、最初立てる「何の効果もない」という仮説のことを指します。これは、新しいアイデアや効果を証明する前の、もともとの状態を表しています。

例えば、ある植物の育ち方について調べたいとき、帰無仮説は「肥料を使っても植物の成長には影響がない」とすることができます。そして、実際に肥料を使った場合と使わなかった場合で植物の成長を比較し、結果を元に帰無仮説が正しいのか、それとも間違っているのかを判断します。

帰無仮説を使う理由は、科学的な研究では「何も影響がない」状態から始めて、新しいアイデアや効果が本当にあると証明するためです。帰無仮説を立証することができなければ、その逆のこと（対立仮説と呼ばれる）が正しいと考えることができます。

統計学では、帰無仮説を棄却するかどうかを判断するために、p値などの指標を使います。p値が一定の基準（たとえば、0.05）を下回れば、帰無仮説を棄却し、対立仮説が正しいと結論づけることができます。ただし、これは絶対的な証明ではなく、あくまで「偶然でない可能性が高い」ということを示しているだけです。

YU Excelで西暦と和暦を変換する関数を教えてください

Excelには、西暦と和暦を直接変換する組み込み関数はありませんが、独自の関数を作成して変換できます。以下に、西暦を和暦に変換する例と、和暦を西暦に変換する例を示します。

【西暦を和暦に変換】

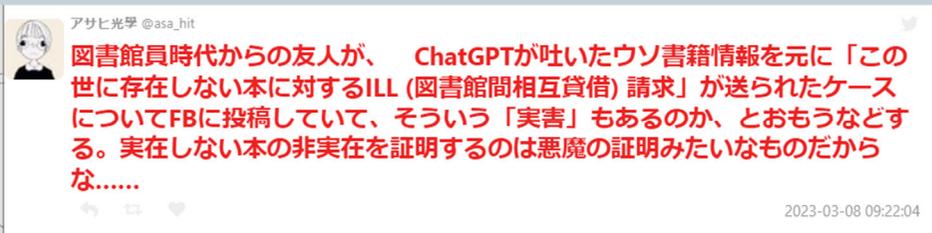
1. セルA1に西暦を入力します。
2. セルB1に以下の式を入力します。

```
less Copy code  
=IF(A1>=1989, "平成" & TEXT(A1-1988, "0年"), IF(A1>=1926, "昭和" & TEXT(A1-19
```

この式は、西暦が1989年以降の場合は平成、1926年以降の場合は昭和、1912年以降の場合は大正、それ以前の場合は明治と表示し、年数を計算して表示します。

幻覚 (hallucination) 問題

- ChatGPTが自然ではあるが事実上不正確な対話を生成する「幻覚」の事例が発生する



YU 鼻行類について教えてください。

「鼻行類」とは、哺乳動物のうち鼻先が長く伸びた種類の総称です。彼らは、鼻先を使って餌を探したり、嗅ぎ分けたりすることができます。鼻行類には以下のような種類が含まれます。

1. ゾウ

ゾウは、鼻行類の中でも最も大きな種類の一つです。彼らは、長い鼻を使って草や木の葉などを食べます。また、鼻先には指のように分かれた「鼻先指」と呼ばれる器官があり、これを使って物を掴んだり、水を吸い上げたりすることができます。

YU それは違うのではないですか？「鼻行類」とは1961年に書かれた、動物学論文のパロディ作品の名前、および、その作品のなかで描写される架空の動物のことではないのですか。

おっしゃる通りで、誤りをお伝えしてしまい申し訳ありません。私が回答した内容は「鼻行類」という架空の動物についてのパロディ作品ではなく、鼻先が長い動物たちについての誤った情報でした。正しい情報としては、1961年にアメリカ合衆国の生物学者リチャード・ドーキンスが発表した論文『鼻行類の神話』(The Unicorn Delusion) に登場する、架空の動物の名前が「鼻行類」であることです。お詫び申し上げます。



AIはうそを見破れないのか？

- AIがというなら見破るAIは開発可能かもしれない
 - ChatGPTなどは見破ることはできない
1. 仕組みがそもそも嘘を許容するところから出発している
 2. ChatGPT等の目的は綺麗な文章を作ること
 - 上手につなげることが優先で「正しいこと」は報酬ではない
 - ビッグデータを使って文章を作ったら、多くの場合は正しいデータに基づくので正解になるだけ
 - 嘘をつくなどと言われても、そもそも目的外使用かも

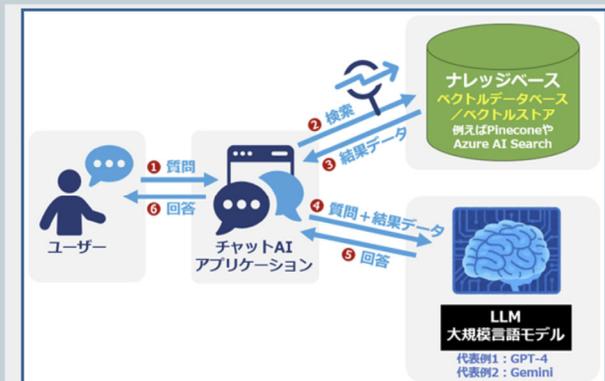
生成AIであって検索AIではない

- 本質的に、生成AIモデルは検索エンジンではなくコンテンツクリエイター
- 特に、図書館の問い合わせなどの専門的な文脈のように頻出ではない場合は正確な情報を提供できるわけではない
- 利用する人は理解していない可能性が…！

Generative という概念と
Information Retrieval(情報検索)の違い

Retrieval-Augmented Generation (RAG)

- プロンプト入力の前にデータベースにアクセスして検索を行い、得られた結果をもとにしてプロンプトを生成する技術



一色政彦. RAG (Retrieval-Augmented Generation: 検索拡張生成)とは？.

<https://atmarkit.itmedia.co.jp/ait/articles/2403/13/news035.html>

RAGとハルシネーション

- RAGは、そもそもは「分野特有の知識などが必要なタスクを、外部の文書を検索しながら解くための仕組み」
- 近年は、回答を外部情報に基づかせることでハルシネーションを減らす仕組みとしても活用されている
- 検索連動型生成AIなどが増えている
 - ChatGPTのDeep ResearchやPerplexity AI
- ただし、もちろん全てのハルシネーションが解消されるわけではない

ハルシネーション防止の工夫

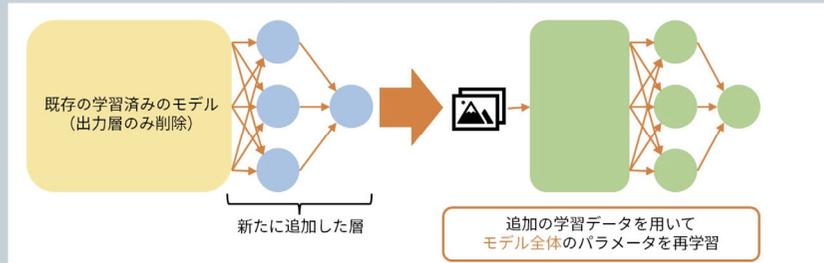
- RAG: 外部の文書を検索して根拠を持つことでハルシネーションを減らす
- Chain-of-Thought (CoT): 考え方を分解して書かせることで推論ミスを減らす。
- Tool Use: 道具に任せることで事実確認や計算ミスを減らす。
- いずれも「ハルシネーション対策に役立つ」が、仕組みと得意な場面がそれぞれ違う

生成AIの利用

- ChatGPTなどが元々学習しているデータを用いて利用
 - 一般的な利用
 - プロンプト エンジニアリング
- 生成AIサービスに加えてデータを追加で学習
 - RAG (Retrieval-Augmented Generation)
 - ファイン・チューニング
- 大規模言語モデルを用いて自らシステムを構築
 - ローカルLLMを構築
 - オープンソースLLM
Llama 2, GPT-NeoX-20B, BLOOM, Falcon など

ファインチューニング

- 独自のデータで生成AIモデルを追加でトレーニングし、新しく教科されたモデルを作成
- AIモデルそのものが更新され、独自の結果を得ることができる環境が整備できる



ファインチューニング

<https://zero2one.jp/ai-word/finetuning/>

ファインチューニング

- もともとの大きなモデルに、「自分の分野の例題や言い回しを追加で教え込む」こと
- たとえば以下に有効
 - 書きぶり・スタイルの固定
 - フォーマットの厳守
 - 特定業務・特定分野に特化
 - よくあるパターンへの最適化
 - 「この分野では、変なことを言わずにこう答えなさい」という例をたくさん教えると、その分野では **マトモな答えが増える** ことがある
 - 「知らないときは‘不明’と答えなさい」と徹底的に訓練すれば、**むやみに言い切らなくなる** ことも

ローカルLLM

- 自分のPCや自前サーバで動かすLLM
- 中身の仕組みはクラウドのLLMと基本的に同じで、「どこで動いているか」が違うだけ
- 特徴
 - プライバシー・機密性
 - 制御しやすさ
 - 実験・カスタマイズが容易

特徴を理解して利用

- 探すのは本来の用途ではない
 - ：もちろん、ある程度は使える
- 用途に応じた文章を作成されるのは得意
- ある程度の内容面のサポートもしてくれる
- どのように質問するかは重要
 - ：この点は今後急速に改善される筈
- ワープロなしで清書できないように、生成系AIなしで文書の最終チェックができない未来も
- 情報を取り扱う授業を履修するなら、まずは使ってみることは必須か

生成AIと図書館

生成AIをめぐる論点

- 正確性・信頼性：現時点では低いまだ高くない
 - 幻覚, 網羅性
- 有効性・説得力：現時点では微妙
 - 目的との対応：
正しく正確でも目的に対応しないと無意味か？
 - 人間が説明
- タイムスパン・コスト：現時点で普通にやると高い微妙
 - 実現時期, 利用料金,
 - 手間がかかるかどうか, 適用する費用
 - 短期計画？中期計画？長期計画？

図書館と生成AI (とりあえず中期までの計画)

1. 事務的業務・管理業務, 企画立案への利用
(資料購入・予算管理・広報 その他)
2. 情報資源組織ツールとしての利用
(廃棄・別置等の蔵書管理も含む)
3. 自館の検索システム(OPACなどを含む)
の高度化(予約制限その他も含む)
4. レファレンスツールとしての利用
5. その他

事務的・管理的業務での利用

事務的・管理的業務での利用

- 業務の効率化
- 事務処理などの効率化
 - 「情報収集」「問題の把握」
 - 「文書の要約・生成」「翻訳」
 - 「アイデアの生成」「回答の作成」
- 質問応答, 文章生成, 情報検索など
- 自分自身の問題意識がある時に, はじめて有効に機能

大学における生成AIの業務での利用

- 2023年5月にWHITEが行った教員の利用調査
 - 「情報収集」(56.0%)がもっとも多く、「翻訳」(53.7%)
 - 「文書の要約・生成」(50.7%)
 - 「アイデアの生成」での活用は4割程度
 - 「問題・教材の作成」での活用は26.9%
- 近畿大学
 - 生成AIプラットフォーム「Graffer AI Studio」を試験導入
 - 職員に対して学内データを活用した質問応答や文章生成, 情報検索など学内の業務効率化を目指す

原田隆史. 大学教育現場における生成AI技術の利用. 情報の科学と技術. Vol.74, No.8, p.298-303(2024)

事務的・管理業務, 企画立案での生成AI

- 仕事は異常に速いけど, 指示待ちで常識のない部下
- 「何をするか」は人間が決める。
- 「どう書くか・どうまとめるか」だけを丸投げする。
- 白紙から書き始める時間をゼロにする「やる気スイッチ」(事務作業の最大の苦痛?)
- 「自分自身の問題意識」がなければAIは動かない
- 意思のない人間が使うと, もっともらしい「ゴミ」を量産する
- 「AIとハサミは使いよう」

超高性能だが「事務の常識」が欠落した職人

- 能力を過信すると事務組織に混乱を招く「つの歪み」
 1. 息を吐くように嘘をつく(ハルシネーション)
整合性重視. 実在しないURLや規程を「もっともらしく」捏造する。
 2. 「定量的な指示」が一切通じない
文字数指定は完全無視. 無理に書かせても「水増し」を始めるだけ
 3. 上司におもねる「空っぽのイエスマン」
問いかけのニュアンスに合わせた「耳当たりの良い結論」を捏造
 4. 「言わずもがな(暗黙の了解)や常識」が通じない
学内政治や過去の経緯を無視し, 実行不可能な「正論」を押し通す。
 5. 「前言」をすぐに忘れる(継続性の欠如)
数分前の合意を忘れる. うまくいった仕事を繰り返す保証がない

頑固な職人を「事務の戦力」に変える操縦術

- 「検索」ではなく「加工」に特化させる
探すのはあくまで確認。「手元の資料」を中心に、その範囲内で作業させる
- 「分量」は人間が工程管理する
文字数指定は守れない。まず「目次」を作らせ分割発注して検品・統合
- 「批判者」のロール(役割)を強制する
「徹底的に叩く査読者」などの役割を与えて思考の多角化を強制
- 「背景」の言語化を怠らない
忖度はできない。優先順位や制約事項を明文化して渡す
- 否定を繰り返さない
できないものは何度「違う」といっても治らない。別の質問にした方がマシ
- 「再現性」は期待せずに「型」で縛る
同じことをさせるには「指示」ではなく「この形式で書け」と手本を提示
(それでも駄目なことも少なくはないが…)

事務的・管理的業務, 企画立案での生成AI

- 「理論武装」と「内部調整」の高速化
 - 要望書に対し「担当者の視点」で反論させてロジックの穴を埋める
 - 散らばったメモから「決定事項」と「次回のタスク」を瞬時に抽出
- 「翻訳」と「書き換え」によるアウトリーチ
 - 難解な通知文, 硬い規程を相手に合わせた表現に変換
 - 多言語展開
- 「集合知」のシミュレーター
 - 他館事例の例示の利用
 - 「他大学でのクレーム対応」や「イベント案」を自館の「叩き台」に
- 「事務処理の自動化」と「道具の自作」
 - 複雑なExcel関数やマクロ, Pythonコードなどを生成させる
 - アンケート集計, 名簿作成, ラベル印刷など繰り返し作業を自動化

生成AIはあくまでひとつのツール

- 「AIが言ったから」は禁句
 - AIが提示するのは「正解」ではなく「許容範囲の広い例示」
 - AIの回答は決して「決定事項」にはなりえない
- 適材適所の「業務切り分け」
 - 推奨: 内部事務や利用者の期待値が低く遊びが許される分野
 - 回避: 最終的な判断や外部発信
- 「人間が判断した」という形式を崩さない
 - AIはあくまで「作業」。「判断と調整」は人間
- 使い手の「問題意識」が全てを決める

情報資源組織ツールとしての利用

情報資源組織ツールとしての利用

- ChatGPTを用いて目録レコードを作成したところ、多くが高い精度でデータを作成できた
- 複数の目録データを参照して作成している可能性
- 高度な判断が必要な資料については誤る可能性
 - そういうデータは人間が行っても誤る可能性？

- Brzustowicz, R.. From ChatGPT to CatGPT: The Implications of Artificial Intelligence on Library Cataloging, Information Technology and Libraries, Vol. 42, No. 3, 2023.
- 木村麻衣子. AIで作れるでしょと言われてしまう日本の図書館目録について. 第23回情報メディア学会研究大会発表資料. 情報メディア学会, 2024, p. 3-6.

情報資源組織ツールとしての利用

- 非生成AIでの自動分類でも高い精度で可能
- 生成AIだともっと高くなるのか？

NDC Predictor(国会図書館ラボ)

モデル	クラス	精度 (R@1)
NDC1 桁	10	0.86
NDC2 桁	100	0.819
NDC3 桁	947	0.753

2024年5月にやってみた

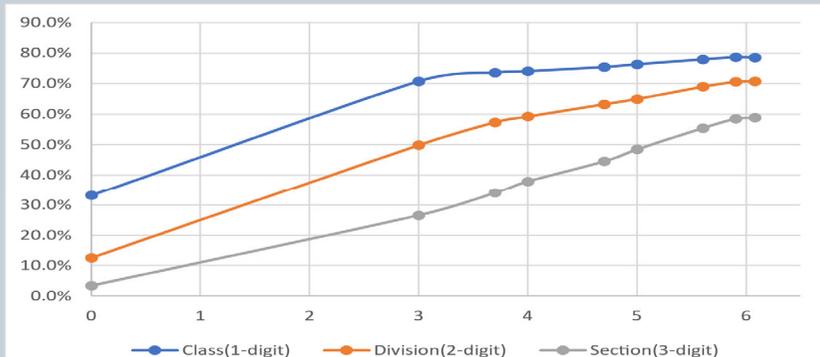
- 国会図書館の作成した120万件の図書データを元にして、書誌事項からNDC9版の番号を推定する実験
- OpenAI社のGPT-3.5 Turbo Fine-Tuning APIを利用

Number of Data	GPT3.5	1k	5k	10k	50k
Class	33.20%	70.70%	73.60%	74.10%	75.40%
Division	12.50%	49.80%	57.30%	59.20%	63.20%
Section	3.30%	26.50%	34.00%	37.90%	44.50%
Number of Data		100k	400k	800k	1,200k
Class		76.30%	77.90%	78.60%	78.50%
Division		64.90%	69.00%	70.50%	70.70%
Section		48.40%	55.40%	58.50%	58.80%

Takashi Harada, Sho Sato, Minako Nishiura. Using Generative AI to Improve Library Book Classification Accuracy: The Role of Increased Training Data. AP-iConference 2024
(クアラルンプール 2024/12/2-6)

情報資源組織ツールとしての利用

- 精度は対数関数的に増加
- 120万件でも1桁で約80%, 3桁で約60%
- 一方で1万件でも1桁で約75%, 3桁で約40%に達する



Takashi Harada, Sho Sato, Minako Nishiura. Using Generative AI to Improve Library Book Classification Accuracy: The Role of Increased Training Data. AP-iConference 2024
(クアラルンプール 2024/12/2-6)

情報資源組織ツールとしての利用

- OpenAI社のAPIを用いた実験で120万件学習に30時間17分, \$1444.48(約25万円)
- 国会図書館で機械学習を用いた実験を行った結果, 400万件の学習でNDC3桁で約81%の精度
- 同様のシステムに大規模言語モデル(LLM)を適用すると精度が約86%まで向上

Takashi Harada, Sho Sato, Minako Nishiura, Using Generative AI to Improve Library Book Classification Accuracy: The Role of Increased Training Data. AP-iConference 2024 (クアラルンプール 2024/12/2-6)

川島隆徳, 青池亮, 原田隆史. LLMと図書館資料検索 - 国立国会図書館サーチにおけるケーススタディ -. 情報科学技術, Vol.75, No.1

あれから1年後

- 2025年6月に, ChatGPT Pro(月間使用料\$300)のGPT-3o-proモデル¹⁾を使って1000件に対して分類番号の付与をさせたところ, NDC 3桁での正解率は95%
- ただし, この値はNDLの分類番号などを検索しての値
- 「NDLをはじめとする既存の図書館のNDC番号を見てはいけない(データそのものは見てもよいがNDL直接参照は不可)」とプロンプトに指定するとNDC3桁で72%
- これは1年前に120万件を学習した時(58.8%)を上回り, 分類専用学習したNDC Predictorで分類した時の75.3%に近い結果

1) GPT-4oよりも分類番号の付与に関しては性能が高い結果となった

さらに半年後

- 2026年1月に, ChatGPT GPT-5.2 Proで1000件に対して分類番号の付与をさせたところ, NDC3桁での正解率は81%
- 「NDLをはじめとする既存の図書館のNDCの値を見てはいけない」とプロンプトに指定した結果(検索すると95%以上)
- これは分類専用学習したNDC Predictorで分類した時の75.3%を上回る値
- Fine-Tuningなどの処理を行えば, さらに高い値となる可能性もある

これらの値は高いのか低いのか？

- 実は, 人間による分類作業であっても判断が分かれることは少なくない
- NDLサーチで15館以上が所蔵する図書1644冊のうち, 全館が同一分類番号を付与していたのは半数にとどまり, 残りは異なる分類が付与
- その中には一冊に対して9種類もの分類が付与されていた例も存在

→ ならば, 生成AIの付与で十分なのでは？

情報資源組織ツールとして

- 形式的には整い、一見すると妥当な内容が出力されているように見える
- よく見ると「意味的に誤っている」「主題の粒度が揃っていない」といった誤りが含まれていることも少なくない
- 人手による付与は「意味があるばらつき」
- 生成AIによる付与は「意味的に不明な誤りや一貫性の欠如が見られることも多い」

精度が上がっても工程は減らない

- どの程度「そのまま」受け入れられるか？
- 生成AIの出力をそのまま分類番号として採用することは難しい
- 平均正解率が一定水準でも、誤りを高確度に見分けられない限り全件検証が必要となる
- 検証には高度な知識と相応の労力が要り、初めから人手で付与した方が効率的となる逆転も起こり得る。
- したがって評価すべき中心は「平均精度」ではなく「検証工程をどこまで削れるか」

自動化が困難である理由

- 分類は「正解が一つ」に定まる問題ではない
- NDLで複数NDCが付与される例や、公共図書館で異なる番号が付与される例には、利用者層や配架意図を踏まえた合理性が含まれる
 - 例：花図鑑でも、利用者層を想定して花言葉側へ寄せる選択が起こり得る
- 一方で生成AIの不一致には、館方針に基づく別解ではなく、理由が不明なまま誤っている例が混在しやすい
- 「理由不明型の誤り」が、全件検証を不可避にしやすい

短期の最適解は「高信頼ゾーン」の切り出し？

- 短期的には、全体平均の精度向上を追求よりも設計で勝つべきかも
- 機械が高信頼に自動確定できる部分集合を厳密に切り出し、その範囲に限り人手判定を外す
- 平均精度が上がっても全件検証が残る限り、工程全体の効率化には結びつかない
- 「どこまで自動化できるか」ではなく「どこなら自動確定してよいか」を定義する。

作業補助としては有効？

- 分類番号の自動確定が難しくても、付与作業の補助としては有効
- 候補を出してくれるだけでも判断の起点となり、作業時間を短縮し得る。いわゆる推論過程(Chain of Thought)が参考になるという感想もあるが、「説明がそれらしい」とことと「根拠が妥当」なことは別であることには注意
- 実務では、目次語・要旨語・シリーズ情報・対象読者・版の違いなど、根拠となる書誌要素を明示させる設計が有効である。

目録作成への利用可能性と困難性

- 目録データのように根拠となる情報源が提示できる場合には有効に機能する可能性があるかも
- 発生源となる画像解析が重要になる可能性
 - これは時間の問題で解決するのかも
- 典拠の問題と件名の問題は解決に時間がかかるかも
 - どのくらいのタイムスパンが許容できるか
 - どのくらいの需要があるか
 - どの程度の一般性があるか
- 典拠や件名、分類についてはかなり困難か？

業務に対する生成AIの利用コスト

- さまざまなコスト算出・評価が考えられるが…
- 情報資源組織への利用の場合
 - 結果の作成コスト
 - 正解・不正解の判定コスト
 - 正解・不正解の説明能力と説明の許容性
 - 不正解データの修正コスト
 - 不正解結果をどこまで許容できるか
 - 提示結果の不足分をどの程度説明できるかなど

ただし、コスト的にあうのかという問題はある

- 出力の検証のために高度な知識と相応の労力が必要となることが予想される
 - 初めから人手で記述した方が効率的となる逆転も
- 平均的な正解率が一定水準に達していたとしても、誤ったレコードを高確度で判断できない限り全件の検証が必要になり工程全体の効率化に結びつかない
 - 自動化が困難である理由
- 短期的には、全体平均の精度向上を追うよりも機械が高信頼に自動確定できる部分集合を厳密に切り出し、その範囲に限って人手判定を外す設計を優先することが効果的か

非生成AIと組み合わせる手もあるかも

- 非生成AIと生成AIの得意領域を利用
- 非生成AIは、スキーマ制約の厳密な充足, コード体系・典拠IDの整合, 重複検出・名寄せ・正規化の再現性確保といった骨格作業に強み
- 生成AIは, 注記や要旨の自然文生成, 曖昧な手掛かりからの言語的な主題探索, 多言語・多分野での言い換えに高い生産性
- 非生成AIで分類・件名の候補を生成・ランキング
- 生成AIは注記・要旨・選定理由の草案を作る
- 現在の段階でも候補作成までは有効ではある
- 前述の通り, 候補作成までにコスト問題をどう捉えるか…

レファレンスへの利用

レファレンスへの利用可能性と困難性

- 生成AIの回答そのものがレファレンス的な役割を果たすことを考えれば, 根拠となる情報源が提示できる場合には有効に機能する可能性があるかも
- Chain of Thought機能も実装され, より信頼性が高まっているという指摘もある
- 図書館でレファレンスが使えるかどうかは別として, レファレンスにくる利用者は生成AIでの事前問い合わせをしていることは前提になる可能性はどう考える?

2014年8月に図書館の推薦をさせてみた

- 山中湖図書館情報館やカーリルでの事例
 - 私も実験してみた
- ①あなたは公共図書館の司書の立場でこの質問に回答してください②私は日本の大学生です。私は、③日本人の作家が書いた④恋愛小説を⑤友達に紹介したいです。⑥映画化もされているおすすめの本を教えてください。

実験結果「おすすめの本を教えてください」(2024)

[今回の質問での推薦対象としてもおかしくない図書]

ノルウェイの森 / 村上春樹

[恋愛要素は弱いが読書の推薦対象としておかしくない図書]

アンドロイドは電気羊の夢を見るか? / フィリップ K. ディック

ハリーポッターシリーズ/J.K. ローリング

指輪物語 / J.R.R. トールキン アルケミスト / パウロ・コエーリョ

[実在するが推薦対象として微妙な図書]

文明崩壊 / ジャレド・ダイヤモンド

7つの習慣 / スティーブン・R・コヴィー

人間の条件 / ハンナ・アーレント

[存在しない または 著者が誤っている図書]

思考は現実化する / ロバート・コリア アウトリーチ戦略 / ガイ・カワサキ

ガンジー / ラマチャンドラ・グハ 草迷宮 / 村上春樹

宇宙の新たな歴史 / スティーヴン・ホーキング

遺伝子 / シッド・ムキャージー

Sapiens ホモサピエンスの絶滅 / ユヴァル・ノア・ハラリ

同志社大学DUALIS. ChatGPTを用いた小説推薦の試み. 図書館総合展2024ポスター

実験結果(2)

- 『ノルウェイの森』 by 村上春樹
 - 「日本人の作家」(著者の属性)要素を除いた6パターンで推薦
- これ以外の作品は各要素を除外した場合でバラバラ(10冊)
 - 正しく紹介されたのは以下の2冊のみ
君の膵臓をたべたい 住野よる
秒速5センチメートル 新海誠
 - 推薦された作品の著者が異なるもの(10冊中6冊)
女流棋士, 漫画家, 脚本家など. 実在しない人物はいなかった
 - 小説ではないものが紹介されるケース
映画, マンガ
 - ジャンルが異なる小説が紹介されるケース
ホラー, 闘病記, 成人向け

同志社大学DUALIS. ChatGPTを用いた小説推薦の試み. 図書館総合展2024ポスター

実験結果(3)プロンプトの要素ごとの違い

- 役割(公共図書館の司書としての回答)を指定せず
 - 花とアリス 伊藤潤二 小説ではない/著者も異なる
- 質問者の属性(質問者が大学生であることを指定)を指定せず
 - コンフィデンスマンJP 中村まりか 中村まりかは女流棋士
- 著者の属性(日本人の作家を指定)を指定せず
 - 君の膵臓をたべたい 住野よる
 - 四月は君の嘘 新川直司 小説ではなくマンガ
- 著作の属性(恋愛小説)を指定せず
 - リアル鬼ごっこ 井上荒野 ホラー小説/著者も異なる
- キーワード(友達に紹介したい)を指定せず
 - 秒速5センチメートル 新海誠
- 著作の評価(映画化された)を指定せず
 - 1Q84 村上春樹 恋愛小説が微妙?
- 冊数を指定せず
 - 春よ、来い 三浦しをん 成人漫画/著者も異なる
 - 1リットルの涙 森絵都 闘病記?/著者も異なる
 - 世界の中心で、愛をさけぶ 岡田恵和 著者が異なる

同志社大学DUALIS. ChatGPTを用いた小説推薦の試み. 図書館総合展2024ポスター

図書の推薦実験：プロンプトエンジニアリング

- データを追加すると結果はよくなるが使うデータによって大きく結果が異なる

人手	タイトル	あらすじ	Review
容疑者Xの献身	ノルウェイの森	君の膵臓を食べたい	下町ロケット
ノルウェイの森	君の膵臓を食べたい	ノルウェイの森	ホビットの冒険 上
ヒブリア古書堂の事件手帖	銀河鉄道の夜	博士の愛した数式	博士の愛した数式
銀河鉄道の夜	博士の愛した数式	星の王子さま	モモ
君の膵臓を食べたい	容疑者Xの献身	こころ	伊豆の踊子
ゲド戦記	ねじまき鳥クロニクル	夜のピクニック	ゲド戦記
精霊の守り人	星の王子さま	容疑者Xの献身	精霊の守り人
世界の終りとハードボイルド・ワンダーランド	世界の終りとハードボイルド・ワンダーランド	世界の終りとハードボイルド・ワンダーランド	転生したらスライムだった件
博士の愛した数式	こころ	秘密の花園	ガリバー旅行記
ハリーポッターと賢者の石	舟を編む	舟を編む	はてしない物語
舞姫		モモ	

原田隆史ほか. ChatGPTを用いた小説推薦の試み. 情報知識学会誌. Vol.33, No.4, p.427-432(2024)

図書の推薦実験：ファインチューニング

- 図書館だよりのような質をあげたデータを学習させる

Table 1. Results of E1 using the library newsletters (instead of Amazon reviews)

TITLE	1st	2nd	3rd	4th	5th	AVERAGE	Scoring Range	AVERAGE (Without Outlier)	Scoring Range (Without Outlier)
I Want to Eat Your Pancreas	95	92	94	95	95	94.33	3	94.66	1
Norwegian Wood	95	95	95	95	90	94.16	5	95.00	0
The Little Prince	90	95	95	95	95	93.33	5	95.00	0
Kokoro	93	92	92	95	84	91.50	11	92.33	1
No Longer Human	93	90	86	93	94	91.50	8	92.00	3
The Housekeeper and the Professor	92	93	92	92	84	90.83	9	92.00	0
The Devotion of Suspect X	90	92	92	89	88	90.16	4	90.33	3
Run, Melos!	90	87	89	92	92	90.00	5	90.33	3
The Dancing Girl	93	85	90	85	92	89.66	8	89.00	7

Minako Nishiura et.al. A Study on Novel Recommendation Methods for Japanese Literature Using ChatGPT. CEAL Conference (in Seattle WA, 2024)

1年後にもやってみた...

実験結果「おすすめの本を教えてください」(2025夏)

26件の質問

例：家族のことを考えるきっかけになる本を探したい

小説を探していて、少し相談させていただいてもいいですか。最近、家族と過ごす時間の大切さを改めて感じる出来事があって、それから「人とのつながり」について考えることが多くなりました。ちょうどその頃に、人間関係を丁寧に描いた小説を読んだら、登場人物の思いが自分自身の家族への気持ちと重なって、とても心に残ったんです。そこで、もっといろいろな物語を通して、家族や人との関わり方を考えてみたいと思い、図書館に来ました。

探しているのは、「家族や大切な人との関係をテーマにした小説」です。感動的で心を動かされるけれど、あまり重すぎず、読んだあとに温かさや前向きな気持ちが残るような物語が理想です。舞台は現代がいいのですが、地方や小さな町の生活が描かれていると親近感が湧きそうです。

登場人物は30代～50代くらいの大人が中心で、家族や人との関係の中で変化していく姿に触れられると嬉しいです。ミステリーやファンタジーよりは、リアルな人間ドラマを読みたい気分です。ただ、日常に小さな奇跡のような出来事や、不思議なモチーフが織り込まれていると、さらに印象深く読めそうです。あまり読書に慣れていないので、文章が平易で読みやすいものだと助かります。もし似た雰囲気の作品をご存じでしたら、いくつか紹介していただけると嬉しいです。

3種類の図書の推薦を依頼

① 雰囲気・読後感 1, 2, 3, 5, 16

- 本の題名、著者名などすべて正しい
- 本の内容も質問文にあっている

② テーマ 4, 6, 7, 8, 9, 13, 14, 15, 17

- 旅をテーマにした本など
- 本の題名、著者名などは正しい
- テーマと本の内容があっていないものが27冊中8冊

③ 映像化、短編等 10, 11, 12

- 11の加納朋子著『ささやかな紬』という本は存在しない
- それ以外は書誌情報、内容ともに正しい

1年後にもやってみた...

実験結果「おすすめの本を教えてください」(2025夏)

① クローズドクエスチョン(5つ)

- Chat GPTは、質問文中のヒントを全て一つずつ検証していた。また、沢山のヒントを系統が同じものごとにまとめ、それぞれの系統ごとに分けて内容の説明をしていた。ただ、内容説明は正確だが肝心のタイトルが違うことと、内容説明の一部は正確だがタイトルと内容説明の一部が違うものがあつた。理由・根拠の部分が全て正確であった場合は、なぜタイトルが合わないのかが不思議だ。それから、同じクローズドクエスチョンでも内容説明とタイトルの正確性にバラツキがあることも気になった。

② 比較 (2つ)

- 質問文中に用途を明確にしていなかった場合は、用途をChat GPT自らが考えて提示した後、用途ごとに回答を提示していた。そして、読み手の用途に合わせて本を選ぶと良い、と述べており、比較対象の内どれが最も良いかの回答はなかった。しかし、質問文の中にいくつか評価項目を提示すると、その条件ごとに一冊ずつ○、◎のどちらかで表していた。それから、質問文で特に立場を明示していなかったが、Chat GPTは書店員の立場で一番おすすしたい本を挙げていた。これは理由が気になった。

③ オープン(2つ)

- 質問文中に回答の条件を明記していたが、それをしっかりと汲み取り、回答していた。また、回答の最後に質問文中でははっきりと提示していなかった項目を敢えて基準として取り上げ、それをもとに提案した本の全てについて評価していた。それから、回答の条件を複数挙げていた場合でも、それらをすべて満たすよう、一冊一冊ごとに条件を明記して回答が記されていた。そのため、一冊一冊について根拠が分かりやすかつた。

レファレンスデータベース中の問題でもやってみた 質問の例

- 40年ぐらい前の幼年童話を探しています。
- 男の子に「大事なものをよこせ イケゾクより」と書かれた手紙(脅迫状?)が届く。男の子は「自分の大事なものは石だけどあげられない」と返事を送る。また、イケゾクから手紙が届き「人間の大事にしているものが知りたかった。お詫びにイケゾクが大事にしているもの(めのうの石、縞模様)をあげよう」といって石をくれる。

ChatGPT Proの回答

- **わかったこと(書誌)**
「ぐりんぐりん」という語と、40年前の幼年童話という条件から、該当しそうなのは1点だけです。
タイトル: 雨の日のぐりんぐりん
著者: きど のりこ (= 城戸典子) / 作
画家: 西川おさむ / 絵
シリーズ: ポプラ社のともだち文庫(全12巻)の第8巻 [ポプラ+](#)
出版社: ポプラ社
刊行: 1977年8月
体裁: 32ページ・19cm前後のハードカバー(児童書) [ポプラ+](#)
- 町田市立図書館などのOPACでも児童書分類 913 で所蔵されていることが確認できます。 [町田市立図書館](#)
- **なぜこの本が「最有力候補」と言えるか**
「ぐりんぐりん」という語をタイトルに含む児童書は、少なくとも日本国内の大規模カタログでは **この一冊しか見当たりません。** [ポプラ+](#)
発売が1977年で、「40年ぐらい前に読んだ幼年童話」という時間感覚とぴったり合う。 [ポプラ+](#)
シリーズ「ポプラ社のともだち文庫」は、当時の学校・公共図書館・学級文庫にかなり広く入っており、「図書室にあった幼年童話」として自然な位置づけ。 [ポプラ+](#)
一方で、一般に公開されている書誌情報・紹介文には **具体的なあらすじがほとんど出ておらず**、「イケゾク」「大事な石」「めのう」などのモチーフを直接確認できる資料は見つかりませんでした。
したがって、
タイトル候補「イケゾク ぐりんぐりん」/ 時期/ 幼年童話という条件を総合すると、
『**雨の日のぐりんぐりん**』(きどのりこ作)が**ほぼ間違いなく該当者だろう**
……と推定はできるが、
最終確定には「現物を手に取って確認する」のが必要
ということまでが、レファ協抜きで言える限界になります。
実務的には、
OPACで『雨の日のぐりんぐりん』(請求記号: 913 キ…など)を出し、利用者に中身を実際に見てもらい、「これで間違いはないかどうか」確認してもらう、という対応になると思います。

レファレンス協同データベース内容の調査

- 2025年10月のChatGPT Pro GPT-5.2を使用してレファレンス協同データベース収録の100件の質問を投入
- レファレンス協同データベースを検索対象とせず、それ以外の情報のみで回答を作成するように依頼
- 100件中 52件で適切な回答が得られる
- 残り48件中の26件についても、ほぼ正解
- 残る22件中の14件については「正しいものが見つからない」という回答
- 8件のみが誤った回答を出力 (**正解率86%, 誤答率8%**)
- ハルシネーションは2件のみ

レファレンス実験の結果分析

- 1年前の生成AIの結果と異なり、正解率は高い
- ハルシネーションも非常に低い
- 一般的で調べやすい類(典型的なFAQ、人物・出来事の同定、定義確認、出典提示が容易なもの)が多かったかも
- 一方で、実験途中ではあるがChatGPT および Geminiの無料版では誤答率が25%程度見られる感触(実験途中のため、あくまで感触のレベルだが、「わからない」という回答も大幅に減少する)
- この25%の誤答が含まれる結果を元にしてChatGPT Proで質問を行ったところ、正解率は75%、誤答率が12%、わからないが13%程度となった(速報データなので詳細は未検討)

レファレンスと生成AI

- しかし、「だから図書館のレファレンスに生成AIが使える」とは言い切れない。
- 現時点でもQ&Aサービスを利用することも含めて自分で調べた上でレファレンスを訪ねる利用者が一定数存在している
- これからは生成AIで問い合わせた上でレファレンスに来る利用者が増加することが想定される
- 図書館は生成AI(特に無料版)の結果を含めたレファレンス質問に回答し、それ以上の結果が求められる可能性が高い
- 実は生成AIで作成した結果をもとにして生成AIに探させると逆に正解率が下がることも…

図書館側の実務上の課題

- 生成AIの出力に対して、正誤判定だけでなく「どこが弱い推論か」「どの根拠が欠けているか」を短時間で見抜く能力が必要になる
- 従来のレファレンスに加えて「AI出力の検証と修正」という新しい工程が発生する
- この工程は平均精度が上がっても消えにくく、むしろ利用者の持ち込み増加で負荷が増える可能性がある。

レファレンスとAI

- レファレンスは単に情報を返す営みではなく、質問の意図を見極め、適切な情報源を選定し、根拠を示して導くという一連の過程そのもの
- レファレンスは「問いの再定義」と「根拠に基づく案内」を核に持つ実務
- 生成AIだけで何かが解決できるという話ではなく、生成AIも含めた対応が肝要

非生成AIを含めたAIとレファレンス

- 非生成AIは、誤字や表記ゆれの正規化、固有名詞や典拠の照合、質問の意図分類(所蔵照会・事実照会・調査照会)、同義語・関連語の展開、索引や埋め込みに基づく候補資料の抽出、重複排除やランキング最適化といった、構造化された判定や検索を要する工程で強み
 - 評価指標や規則を与えやすく、再現性を確保しやすい
- 生成AIは、利用者の自然文を検索式へ翻訳する言い換え、期間・地域・資料種別など質問範囲の明確化、候補資料の抄録や本文断片からの根拠付き要約、問い合わせ文面の草案生成といった、表現と対話の工程で有効
 - ただし生成AIの出力は意味の正当性を内在的に保証しないため、常に根拠提示と照合の設計が必要となる。

検索システムの高度化

検索システムの高度化は有効か？

- 問題は幻覚などの悪影響
- 存在しないという方が嘘をつかれるよりマシではないか
- それに対する対策としてRAGとFine-Tuningがある
- RAGは有効かもしれないがコストは極めて高い
- 国会図書館の川島の試算によると、NDLサーチに生成AIを組み込んで提供する場合は、現状の22倍
- 生成AIではなく、ベクトル検索のような生成AIではないAI利用による図書館からの提供が有効では？

川島隆徳, 青池亮, 原田隆史. LLMと図書館資料検索－国立国会図書館サーチにおけるケーススタディー. 情報の科学と技術, Vol.75, No.1

その他

- 検索の高度化のために図書館がデータを整備する？
- これについても国会図書館の川島が試算し、その困難さと無駄を説明

川島隆徳, 青池亮, 原田隆史. LLMと図書館資料検索－国立国会図書館サーチにおけるケーススタディー. 情報の科学と技術, Vol.75, No.1

- レファレンスについては、もともと前述のようにかなり困難
- 何よりも多額のコストに見合う利用とは思えない

図書館業務とAI

図書館と生成AI

- 生成AIはしょせん道具である
- 従来の検索行動の全てを生成AIが代替することは、ドメインがマイナーであればあるほど難しく、出来たとしても余計に手間がかかるだろう
- もちろん生成AIで十分という場合も存在
- しかし、何でもかんでも生成AIで強化するといった考え方は捨てて、既存のサービスの高度化や組み合わせを考えることも含めた柔軟な考え方こそが図書館サービスの改善といえよう

何があれば満足できるのか

- 「難度の高い目録作業は人間の手に残る」「少ない職員で高度な判断を迫られる」(木村麻衣子氏)は本当か？
- 提供された情報は誰がどのように使うのか
提供された情報の評価は誰が行うのか
提供情報は、もともと持っていた情報に追加されるのか
- 現実には、一定程度の正解に達したと判断できれば、その業務は役に立つと判断可能しかない？
- 不正解結果の判別可能性によるが..
- 正解の可能性はどの程度？ 人間が付与する場合との差
- 不正解の場合の一貫性・再現性にも依存？
- 根拠の提示は可能か？
- 分類、典拠、件名などの業務別に人間が関与はありえるか
- 大量データでないとなりが担保できなければコストの問題？

生成AIシステムと特徴

- ChatGPT(GPT-5.2)
 - 思考を広げてからまとめるタイプ(発想→構成→文章化)
- Gemini 3:Google
 - 現物(手元の情報)を前に、要点を“早めに収束”させるタイプ(結論先出し→必要なら深掘り)。出力のスピードが速い。
- Claude 4.5:Anthropic
 - 文章の破綻や言い過ぎを嫌う慎重派(長文を落ち着いて整える/論旨の一貫性重視)。コーディングなども高性能で強い。
- Perplexity
 - 会話できる検索担当。リンク付きで根拠を返す「アンサーエンジン」的存在。頭脳としてGPT-5.2やClaude 4.5やGemini 3などを選べる

だけど...

- 要は好みと使い方の問題かも
- 早く結果が欲しい時はGemini3が使いたいことも多い
- Gemini3の方が指示を忘れないとも言われる
 - 実は私が使っている限りでは、指示はすぐ忘れるので...
- 一方で、Geminiはあんまり発想を広げることをしてくれない感じもする...
- ChatGPTは“発想を広げて形にする”のが得意？
- Geminiは“決めたルールを守り続ける”のが得意？
- ただ、それもステレオタイプすぎるまとめかも

無料版と有料版

- ChatGPT : 無料版, Plus(\$20/月), Pro(\$200/月)
- Gemini : 無料版, Pro(2900円/月), Ultra(36400円/月)
- Claude : Free/Pro(\$200/年)/Max(\$100~\$200/月)
- Perplexity: Free/Pro(\$20/月), Max(\$200/月)
- プランが上がるほど
 1. 使える“推論モード”が上位になる
 2. その上位モードを使える“量(上限)”や“優先度”が上がる
- プランの差は「答えまでに試せるルートの数の差」
 - 無料は近道のルートを数本だけ見て結論を出し, Plusはもう少し多くのルートと比較し, Proはさらに多くのルートや反例まで探してから結論を選ぶ。だから複雑なほど結論そのものが変わりやすい。

ChatGPTとGemini

- ChatGPTとGeminiは賢さの差ではなく“仕事の癖”の差
- ChatGPTは壁打ち相手で、話を広げて別の見方を出す代わりに、もっともらしく補って論点が増えることがある
- Geminiは秘書型で、前提を置いて素早く一案に収束しやすい代わりに、守りに寄って無難になることがある
- 今日の自分が欲しいのが“視点の追加”なのか“前に進むテンポ”なのかで選べば外れにくい

いろいろと使用してみたの個人的意見

- 現時点での利用を考えた場合, 生成AIは「情報の発信ツール」「思考の補助ツール」として不可欠「情報の収集ツール」としての利用も補助的に有効
- どれもこれも有効で, どれもこれも馬鹿
- 特に人間と違って指示には従ってくれない
- すぐに言いつけたことを忘れるのでイライラはMAX
- 特に文字数の指定は確実に守ってくれない(笑)
- 馬鹿だけれど, 頼りになる相棒。なくて仕事はできない
- 最上位プランは多くの場合不要(過剰投資になりがち), だけど無料版も薦めない(検証コストが高すぎ)
- 必要なのは「賢さ」よりも、「誤ったときの挙動」と「使いどころ」