

# AIによる図書館サービスの可能性と課題

国立大学法人電気通信大学副学長  
広報センター長／UECコミュニケーションミュージアム館長  
大学院情報理工学研究科／人工知能先端研究センター教授

坂本真樹

1950年代 - 誕生と初期の研究:1956年: ダートマス会議で「人工知能 (AI)」という概念が誕生。  
【第1次AIブーム: 1960年代】 チェスなどのゲームを対象に、探索的に解ける問題などに焦点。  
1970年代 - 人間社会の複雑な問題は解けないとされ冬の時代へ

【第2次AIブーム: 1980年代】 エキスパートシステムの普及:日本の「第五世代コンピュータプロジェクト」などにより再び注目が集まる。スタンフォード大学の「マイシン」など医療系エキスパートシステムが開発され、感染症診断に応用が試みられた。  
手動で知識をコンピュータに入力することの限界などにより、資金も不足し、再び冬の時代へ

【第3次AIブーム: 2000年代】 1990年代 - コンピュータの性能向上、インターネットの普及により、大量のデータをコンピュータに入力することが容易に⇒2010年代 - ビッグデータとディープラーニング: ディープラーニング (深層学習) が注目を集める。深層学習が飛躍的に進歩し、画像認識や音声認識など様々な分野で実用化。

【第4次AIブーム: 2020年代】 AIのさらなる進化と社会への影響:GPT-3などの大規模言語モデルが登場し、自然言語処理の分野で大きな進展。AIは医療、自動運転、金融など多くの分野で社会的インパクト

## 2012年画像認識AIの急進化

ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2012

1000クラスの物体画像識別コンペティション

Deep Learning（ディープラーニング：深層学習）を用いたシステムが圧勝  
毎年1～2%ずつ向上していた識別率が一気に10%向上

### 【画像認識AIの応用事例】

- 医療分野: 病変の検出: X線やMRI画像からガンやその他の病変を自動的に検出。診断支援: 医師の診断をサポートするためのツールとして利用。
- 自動運転: 道路標識の認識（自動車が道路標識や信号を認識）、障害物検出（他の車両や歩行者をリアルタイムで検出）⇒安全運転を支援
- セキュリティ: 顔認識（セキュリティカメラでの個人識別やアクセス制御、空港でも実利用）、行動解析（カメラ映像から不審な行動を検出）
- 小売業: 商品認識（商品棚の在庫管理やセルフレジでのバーコードレス商品認識）、顧客分析（店内カメラで顧客の行動を分析し、マーケティング）
- 農業: ドローンやカメラを使って作物の健康状態をモニタリング、画像認識を使って害虫の発生を早期に発見など

コンピュータの生みの親と言われる  
イギリスの数学者のアラン・チューリング（1912年～1954年）  
人工知能の出来栄えを判定するチューリング・テストを考案

## ヒトと共存するAI搭載ロボットに ほしい基本的能力

「見る」能力	➡	100点
「聞く」能力	➡	80点
「対話する」能力	➡	50点
「共感する」能力	➡	50点

と言われてきました

## 2022年11月 対話型生成AI ChatGPT登場

深層学習アルゴリズムの一つである  
GPT (Generative Pre-trained  
Transformer) -3.5ベース=トランスフォーマー  
と呼ばれるモデルアーキテクチャを採用  
トランスフォーマー：長い文章を扱う場合に非常に効果的なモデル  
⇒大量の自然言語データを学習することで、  
AIによる自然な対話を実現  
与えられた入力文を理解し、それに適した返答文を生成するための条件付き言語モデルとして  
訓練

特徴①単語はベクトル化される  
⇒近い意味の単語は距離の近いベクトルになる。  
例えば、「みかん」と「りんご」は近いが  
「みかん」と「エアコン」は遠い

りんご → (0.2, -0.1, 0.55, ....)  
みかん → (-0.1, 0.31, 0.2, ....)  
犬 → (0.4, 0.01, 0.8, ....)  
エアコン → (0.71, 0.67, 0.31, ....)  
ノート → (0.314, 0.83, 0.48, ....)  
...

特徴②AIは**コーパス**から、自然な文章になるような単語の並び方を学習する。  
・文章生成は、スコアが高くなるように単語をつなげる。  
例) 「○ ○を食べる」のような穴埋めする問題をひたすらやる  
「りんごを食べる」と「みかんを食べる」はほぼ同じくらい高スコアになる。  
「エアコンを食べる」は低スコアになる。  
特徴③AIによる文章生成は意味をつなげるのでコーパスにない文も生成される。

コーパスが大規模なほど精度が上がるため、  
大規模言語モデル (LLM : Large Language Model) 競争へ  
例) データ量 : GPT-3は約3,000億パラメータ⇒GPT 4 は非公開だが兆単位

## 【Step 1】 教師あり学習

入力（プロンプト）は、  
アノテータがGPT-3から  
OpenAIのAPIで収集

その内容をラベリング  
プロンプトに対して  
適切なレスポンスを記述  
各入力に対して最適な  
出力を生成

新しい教師あり  
データセットでGPT-3を  
微調整  
GPT-3.5 (SFTモデル)  
を作成

## 【Step2】 報酬モデル (RM) 学習

報酬を最大化する出力を生成する  
ように学習する  
強化学習で利用するために必要な  
モデル

ラベル付け担当者には、  
1つの入力プロンプトに対する  
4~9個のSFTモデル出力が与え  
られる

ラベル付け担当者は、  
これらの出力をベストからワース  
トにランク付けするよう求められ  
、出力ランクの組み合わせを作成

## 【Step3】 強化学習

モデルにランダムな  
プロンプトが提示され、  
応答が返される  
このレスポンスは、ステッ  
プ2で学習した方針（報酬を  
最大化するモデル）で生成

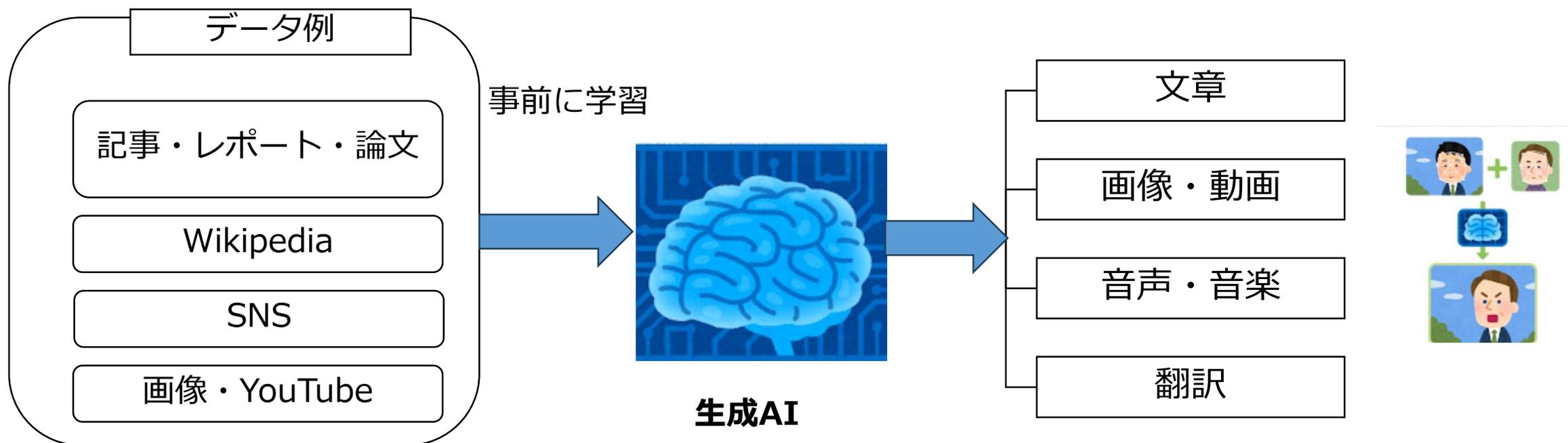
報酬はモデルにフィードバ  
ックを行い、  
方針を進化させる

Schulmanら (2017)  
Proximal Policy  
Optimization (PPO)

<https://arxiv.org/pdf/2203.02155.pdf> のp.3 Figure2を参考に日本語で説明

- 生成AI (Generative AI) : テキスト、画像・動画、音声などさまざまなコンテンツを生み出す
- 多種多様なコンテンツを作れることから、社会の多様な場面で活用される機会・可能性が増加

- オンライン上にある一連のデータ (データセット) を事前に学習したAIが、利用者から質問や条件の指示 (プロンプト) を受けて、新たなコンテンツを生成



ChatGPT4oに

「生成AIが図書館のサービスでできることを、できるだけたくさん、詳しく、教えてください」と聞いてみました。



## ChatGPTの回答

生成AI（生成型人工知能）は、図書館のサービスを向上させるために多くの方法で活用できます。以下に、その具体的な方法を詳しく説明します。

1. レファレンスサービス質問応答: 利用者からの質問に対して、AIが迅速かつ正確に回答を提供します。
2. 資料検索: 特定のトピックや質問に関連する書籍、論文、記事などを検索し、利用者に提示します。
3. コンテンツ生成要約生成: 長い文章や論文の要約を自動的に生成し、利用者が短時間で内容を把握できるようにします。
4. レポート作成支援: 特定のテーマに基づいたレポートや論文の構成案や初稿を生成します。
5. 自然言語処理テキストの翻訳: 多言語対応の図書館サービスとして、利用者の母国語にテキストを翻訳します。
6. テキスト解析: 文書の感情分析やテーマ抽出を行い、利用者に有用な情報を提供します。

## ChatGPTの回答（つづき）

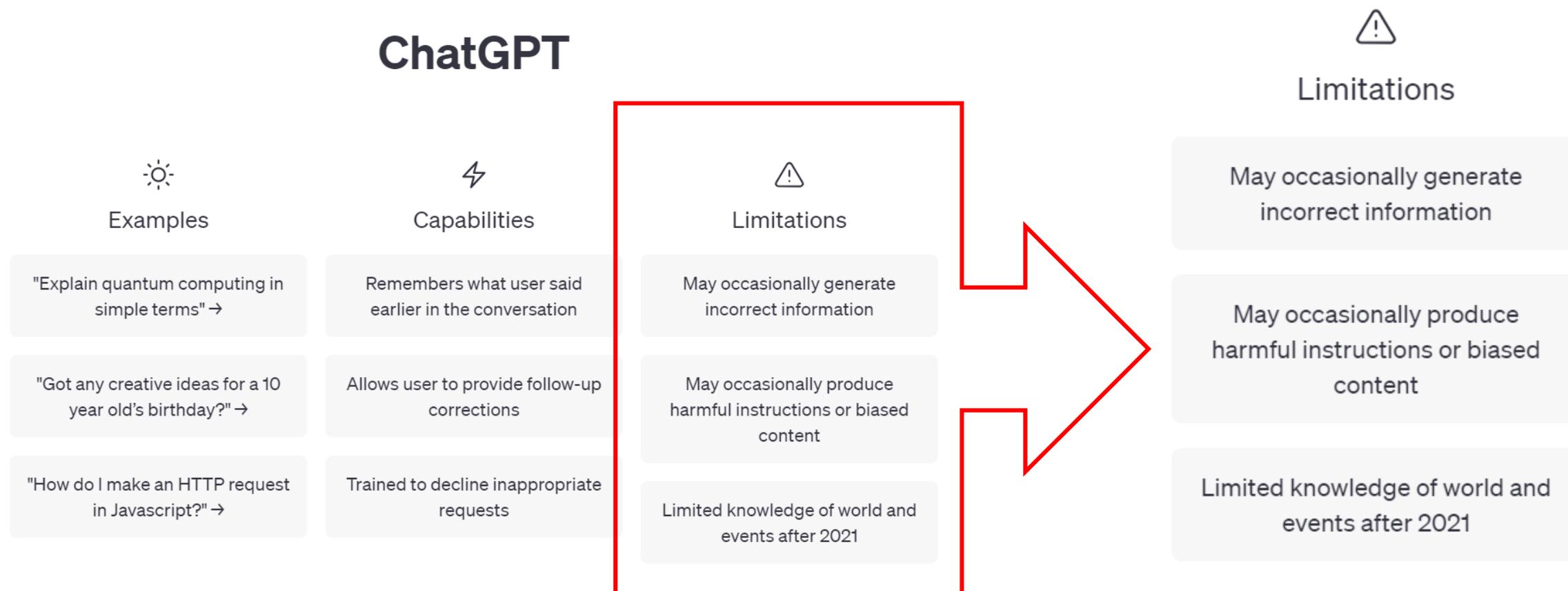
7. パーソナライズドサービス・おすすめ機能: 利用者の過去の借りた本や検索履歴を元に、興味のあるような書籍や資料を推薦します。
8. カスタマイズされた通知: 利用者が興味を持つ新着資料やイベント情報を通知します。
9. デジタルアーカイブの活用・アーカイブの検索と分類: デジタルアーカイブ内の資料を効率的に検索・分類し、利用者に提供します。
10. デジタル展示会: AIを活用して、デジタル展示会のコンテンツを生成し、利用者に新しい形での体験を提供します。
11. チャットボット24時間対応: 図書館の営業時間外でも、AIチャットボットが利用者の質問に回答
12. 自己学習: 利用者の質問パターンを学習し、回答の精度と質を向上させます。

## ChatGPTの回答（つづき）

13. 教育支援学習サポート: 学習に役立つ資料の提供や、勉強方法に関するアドバイスを行います。
14. オンラインクラス: AIが講師となり、オンラインでの教育コンテンツを提供します。
15. 図書館業務の効率化蔵書管理: 蔵書の整理や管理を自動化し、効率化を図ります。
16. 利用者データの分析: 利用者の利用パターンを分析し、サービス改善に役立てます。

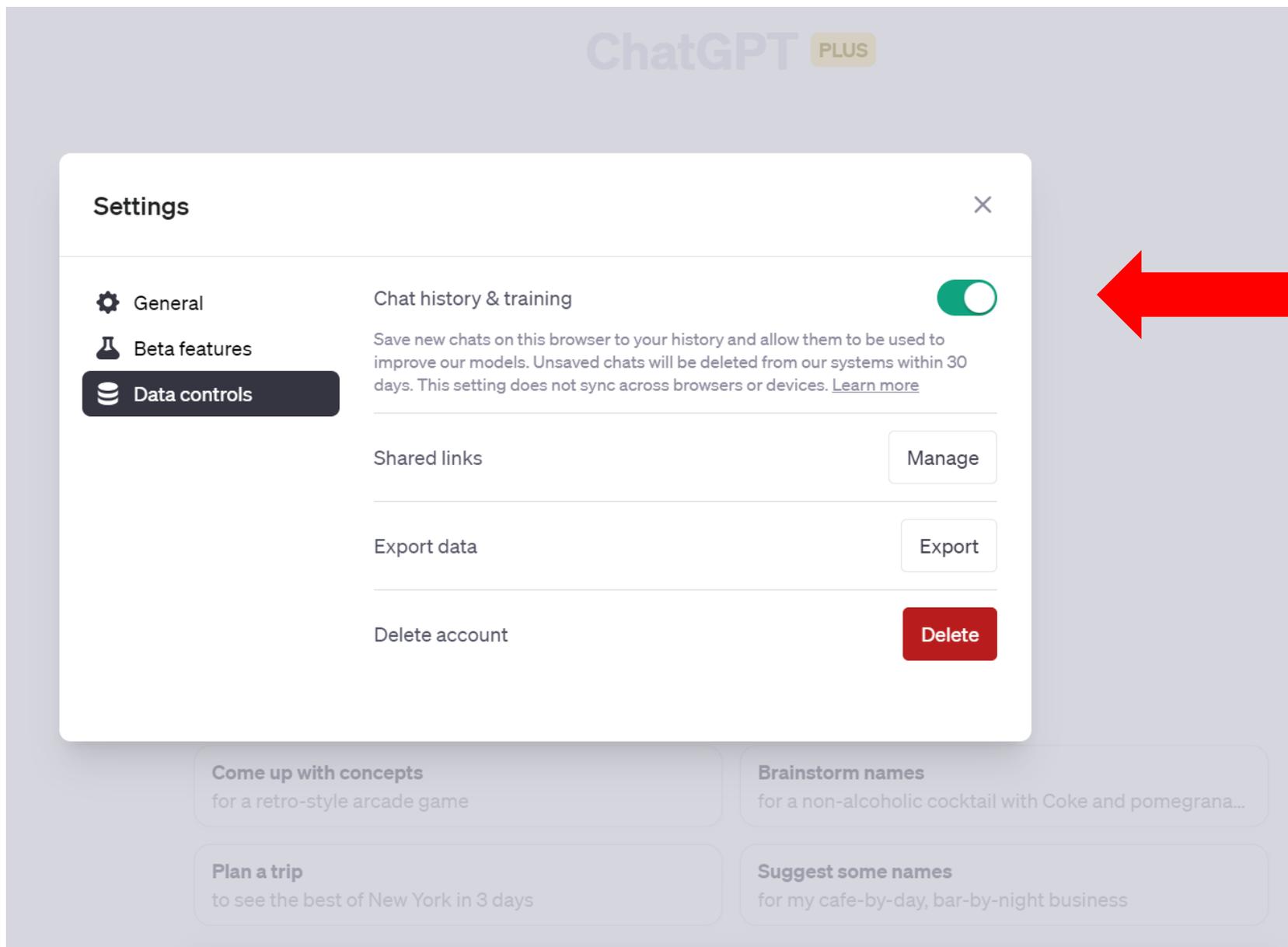
生成AIは、図書館の多岐にわたるサービスを支援し、利用者にとってより便利で快適な環境を提供することができます。これにより、図書館はより多くの人々に知識と情報を提供する場としての役割を強化できます。

生成AIの限界について免責：生成された発言・データに、誤った発言、名誉毀損等になるような発言・不適切な差別発言等が含まれている可能性など  
⇒利用者側の責任：現場教育で生成AIの限界を周知する必要  
⇒外部向けサービスの場合：生成結果について法的・倫理的責任等を問われ得る



## ChatGPTの回答をもとに要点まとめ

- ① 個人情報保護・セキュリティ: 利用者の検索履歴や個人情報など保護対策。データ漏洩、不正アクセスのリスク対策
- ② 利用者の信頼獲得: 新しい技術に対する抵抗感や不信感がある利用者もいるため、AIサービスの受け入れを促進するための教育や説明が必要です。
- ③ 導入コストと運用初期投資: 導入への初期投資。ハードウェア、ソフトウェア、トレーニングなど。  
メンテナンス: システムの運用とメンテナンスのための継続的コスト。
- ④ 偏りや差別バイアスの差別バイアスの問題: AIがトレーニングされるデータに偏りがあると、生成されるコンテンツや応答に偏りや差別が含まれる可能性がある。
- ⑤ コンテンツの正確性・誤情報のリスク: AIが生成する情報や回答が正確である保証が必要。誤った情報が提供されると、利用者に混乱や不利益をもたらす可能性。
- ⑥ 法的および倫理的問題・著作権: AIが生成するコンテンツの著作権に関する問題が発生する可能性。どの範囲までが合法であるかを明確にする必要あり。AIの利用に伴う倫理的な問題の検討。



学習に使われたくない場合は、最低限ここをオフしておく



さらに！  
ChatGPTを利用した  
組織内AI開発  
⇒特定組織内で、セキュリティ環境下でデータを二次利用されずに活用可能

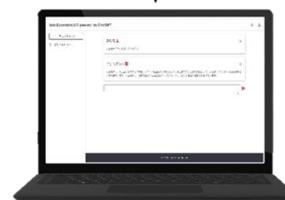
## 生成AIモデルを安全に最大限活用可能な環境構築 (生成AI利用について、非エンジニア・関係者への教育が重要)

	ChatGPTの場合	組織・社内GPT
データ二次利用	原則二次利用される	二次利用されない
データ保管	アメリカに保管	外部に保管されない
データ監視	実施される	実施されない
アクセス制限	だれでも利用できる	利用が認められた人のみ

### 【事例】 Keio Expressive I/O powered by ChatGPT

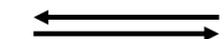
利用者へ活用方法やルールの周知を徹底  
教育コンテンツの充実 (e-ラーニング・  
利用ガイドライン)

ユーザーの約40%  
が 2時間以上/月の  
業務効率化を実感



**Kansei AI!**  
開発/運用  
(アップデート)

サービス利用



グループ全体へ  
サービス提供



\*グループ企業例

Q:電気通信大学について教えてください。

ChatGPTの回答: 電気通信大学(でんきつうしんだいがく、英: The University of Electro-Communications、略称: UEC)は、日本の国立大学で、東京都千代田区(×千代田区 ○調布市)に位置しています。

! 注 ! 2023年5月時点での回答。現在はインターネット検索機能との連携により改善されています。

### ファクトチェックをするサービス開発 (= 画像生成AIなど生成系AIでも需要が大きい)

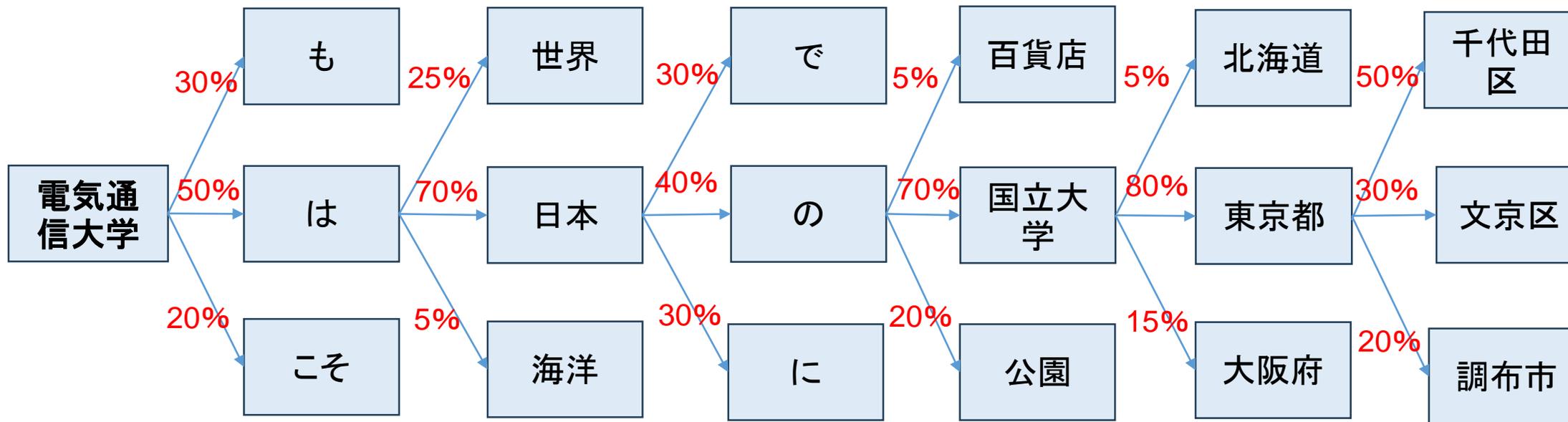
**ChatGPT:** ChatGPTの回答を事実確認するためには、以下のようなサービスを利用することができます :

1. **ファクトチェックのウェブサイト** : Snopes、PolitiFact、FactCheck.org、またはReuters Fact Checkなどのファクトチェックの専門ウェブサイトを利用して、情報の正確さを検証することができます。これらのサイトでは、さまざまな主張や文言の正確性を検証しています。
2. **公式情報源** : 政府の公式ウェブサイト、学術機関、信頼性のあるニュース機関、科学的なジャーナルなどの公式情報源から情報を入手します。これらの情報源では、信頼性の高い情報や研究結果が提供されることが多いです。
3. **検索エンジン** : Google、Bing、DuckDuckGoなどの検索エンジンを利用して、複数の信頼性のある情報源が提供しているかどうかを確認します。
4. **専門家の意見** : 関連する分野の専門家に相談することで、情報の妥当性を確認することができます。専門家は、情報の検証や反証を行うための洞察や専門知識を提供してくれます。
5. **クロスチェック** : 複数の信頼性のある情報源を比較し、情報が一致しているかどうかを確認します。異なる情報源で情報が一致している場合、より事実に近い可能性があります。
6. **図書館の資料** : 地元の図書館やオンラインのデータベースを利用して、学術論文や書籍などの信頼性の高い情報源にアクセスし、情報の検証を行うことができます。

## ④ 偏り・差別 ⑤ 誤情報生成のリスク

Q:電気通信大学について教えてください。

ChatGPTの回答:電気通信大学(でんきつうしんだいがく、英: The University of Electro-Communications、略称: UEC)は、日本の国立大学で、東京都千代田区(×千代田区 ○調布市)に位置しています。



プロンプトエンジニアリング【制約の追加】：プロンプトに特定の条件や制約を追加して、モデルの出力を制限。たとえば、特定のドメインに関する情報をプロンプトに含めることで、そのドメインに関連した回答を得ることができる。

Q:東京駅までの行き方を教えてください。(2023年5月14日時点)

ChatGPTの回答:東京駅までのルート案内をお手伝いいたします。ただし、お答えする情報は**2021年9月までのものであり、現在の最新の情報とは異なる可能性があります。**ご了承ください。

東京駅へのアクセス方法は複数ありますが、以下のような一般的な方法があります。

1. 電車を利用する場合:

- まず、ご出発地点を教えてくださいと具体的なルートを案内できます。最寄りの駅や路線を教えてください。
- 多くの場合、都心部からはJR線や地下鉄を利用してアクセスすることが一般的です。例えば、山手線や東京メトロ丸ノ内線などを利用することができます。
- また、空港から東京駅に直接アクセスする特急列車も運行されています。成田空港からは「成田エクスプレス」、羽田空港からは「東京モノレール」や「京急線」を利用することができます。

2. タクシーを利用する場合:

- タクシーを利用する際には、出発地点の住所や地名を伝えて、運転手に東京駅までの案内をお願いします。

**データのカットオフ問題**＝特定の時点や基準を境にして、それ以降のものを取り扱わない、もしくはそれ以前のものを排除すること。多くの分野や状況で使われるが、訓練データの終了点、つまり学習データが取り入れられた最終の時点。

\* GPT-4の新バージョンGPT4oでは、データカットオフ日(情報の最終収集日)が2023年10月になった⇒最近の出来事についての回答がより正確になった

1. データ品質の向上：AIモデルが学習するデータの質を高めることが基本。信頼性の高い、正確なデータを使うことで、モデルが不適切な情報・誤った情報を学習するリスクを減らす。
2. トレーニング：人間のトレーナーが正しい情報を教える。AIの出力後に、さらにチェックや修正を加える段階を設ける方法（ポストプロセッシング）。人間が最終的な出力を確認し、必要に応じて修正することで、誤った情報の提供を防ぐ。異なるデータセットや異なるアルゴリズムを組み合わせることで、モデルの性能を向上させる。
3. フィードバックループ：  
ユーザーからのフィードバックを活用して、AIの出力を常に監視・改善。誤った情報が検出された場合、すぐに修正を行う仕組みを検討。
4. 透明性と説明可能性の向上：  
AIがどのようにして特定の出力を生成したのかを説明できるようにすることで、誤った情報が出力された理由を把握、問題の特定と修正を検討。
5. RAG (Retrieval-Augmented Generation) 「リトリバル強化生成」：  
リトリバル（検索）：信頼できるデータから、リアルタイムに関連する情報を検索。  
生成:検索して得られた情報をもとに、新しいテキストを生成。

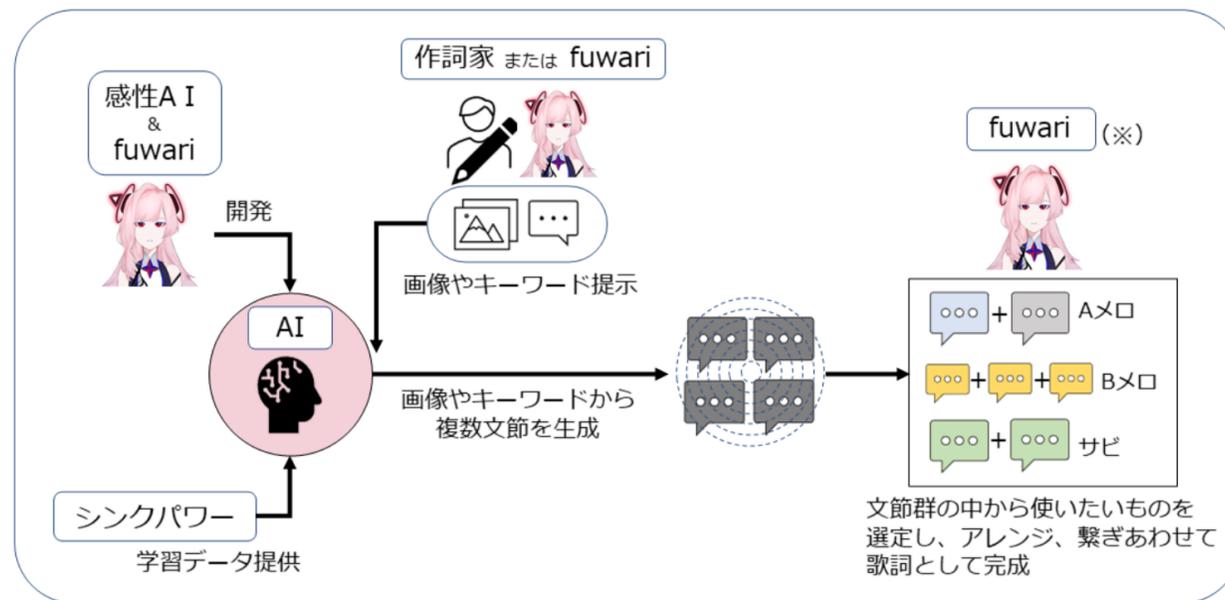
## ⑥ 法的・倫理的問題、著作権侵害のリスクについて

### ● 著作権処理されたデータによる生成

坂本がプロデュースする  
AI作詞家VTuber「fuwari」の取組例

右図の出典

<https://syncpower.jp/ja/news/2022/0330/1>  
(2024年6月23日アクセス)



(※) …歌詞のまとめ上げは作詞家側で実施するケースもあり

### ● AIによる文章や画像の生成の課題をめぐり、日本企業によるサービスも増加

企業事例	時期	概要
NABLAS (東大発ベンチャー)	24年3月	テキスト情報の <b>ファクトチェックツール</b> を提供開始。生成された文章のハルシネーションに対応
	4月	生成AIのフェイクを見破る、企業向けフェイク検出サービスをリリース。
ユーザーローカル	1月	生成AIが書いた文章かどうか見破る「 <b>生成AIチェッカー</b> 」無償提供開始
SnackTime	1月	クリエイターのイラストをAI学習から保護するサービス「 <b>emamori</b> 」リリース

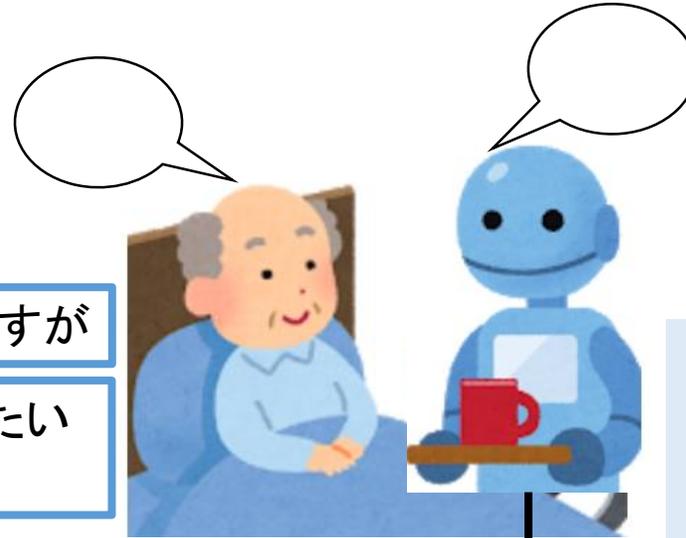
## Cyber空間と直接つながるAIの能力は重要

このあいだ駅前の薬局で買ってき  
た薬をキッチンからとってきておくれ

パンがいい感じに焼けてるか見て  
きてくれるかい？

いつものお医者さんの予約をしたいんですが

図書館でChatGPTについての本を借りたい  
のですが



ChatGPT:申し訳ありませんが、私は物理的  
な場所にアクセスすることはできません。

ChatGPT:申し訳ありませんが、私は物理的  
な場所にアクセスできないため、パンが焼け  
ているかどうかを確認することはできません。

ChatGPT:お医者さんの予約をするには、  
いくつかの方法があります。以下に一般的な  
手順を示しますが、具体的にはお使いの  
クリニックや病院のウェブサイトを確認するか、  
直接問い合わせてください。

ChatGPT:4件のサイトを検索しました。  
...

**注！現在は実際に借りる手続きはできない**

