対話型生成 AI ユーザーマニュアル

富士通株式会社

2023-12-15

| バージョン | 日付 | 変更内容 |
|-------|------------|--------------------------------|
| 1.0 | 2023/9/28 | 初版作成 |
| 2.0 | 2023/11/10 | 幻覚検出機能のアップデート、chat setting の追加 |
| 3.0 | 2023/11/23 | Chat 作成時に利用モデルを選択可能にした |
| 4.0 | 2023/12/15 | フィッシング URL 検出機能の追加 |

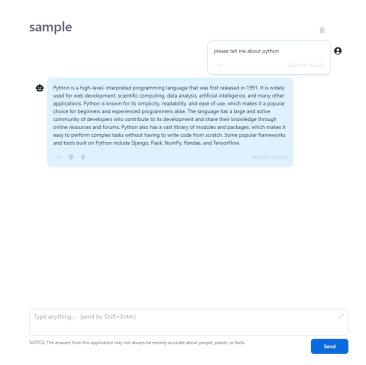
目次

- 1. アプリケーション概要
- 2. ログイン
- 3. ドキュメント管理機能
- 4. 新しいチャットの作成
- 5. チャット
- 6. 幻覚検出機能
- **7.** フィッシング **URL** 検出機能
- 8. トラブルシューティングと FAQ
- 9. アプリケーションの終了

1.アプリケーション概要

このアプリケーションは、チャット形式で生成 AI を試行するためのアプリケーションです。 Azure OpenAI Service の ChatCompletion API を使用し、チャット形式で AI と会話することが出来ます。 また、ドキュメントのアップロードや外部の Web ページを登録し、その情報を参照した形で回答を作成することが出来ます。 FAQ の高度化、要約、翻訳、など様々な用途が生成 AI にはあります。外部データを用いて、どのようなことが実現できるかを本アプリケーションを通して試すことが出来ます。 また、幻覚検出機能を使うと、対話 AI が回答した文章中の「幻覚(データに基づかないもっともらしい誤り)」を検出できます。





利用 AI モデル

- Azure Open Al
 - gpt-35-turbo
 - gpt-35-turbo-16k
- Amazon Bedrock
 - Claude Instant v1

ドキュメント参照機能には以下の AI モデルを利用しています

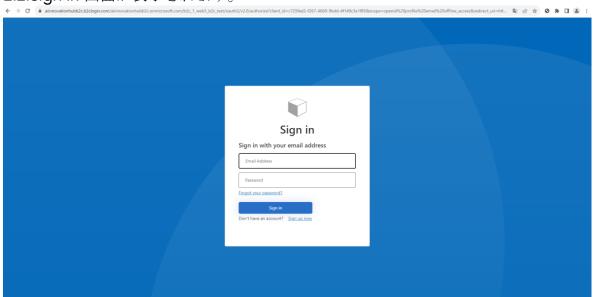
- Azure Open Al
 - text-embedding-ada-002 (version2)

2.ログイン

はじめに、ログイン画面からログインを行います。

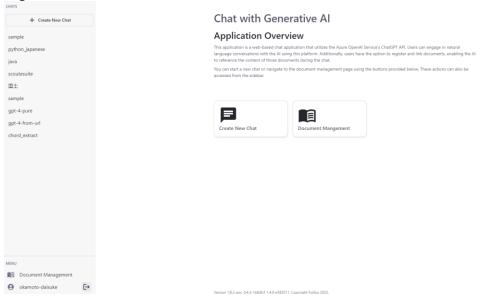


2.2.Sign in 画面が表示されます。

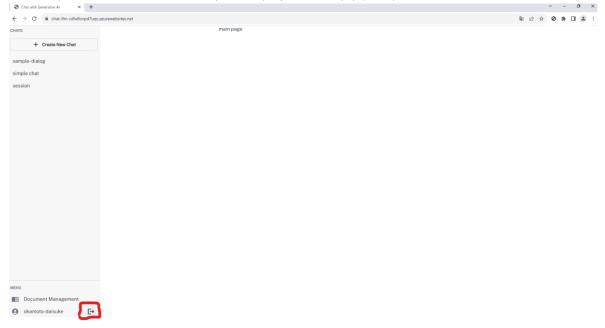


2.3.ポータル画面で作成したアカウントを利用し、Sign in を行ってください

Sign in に成功した場合、アプリケーション画面が表示されます。



2.4.左下のボタンからログアウトを行うことが出来ます



3.ドキュメント管理機能

参照用のドキュメント登録を行う機能を説明します。 登録されたドキュメントの 内容は、チャット作成時に設定することで、その内容を参照した回答を AI にさせ ることが出来ます。 ドキュメントは、一つの情報ソースからなる「ドキュメント」と、ドキュメントを複数組み合わせた「コレクション」が作成可能です。

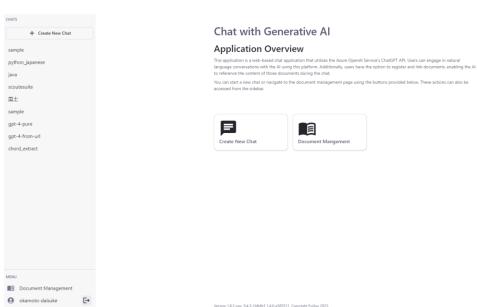
利用可能なファイルフォーマット

以下のファイルフォーマットが利用可能です

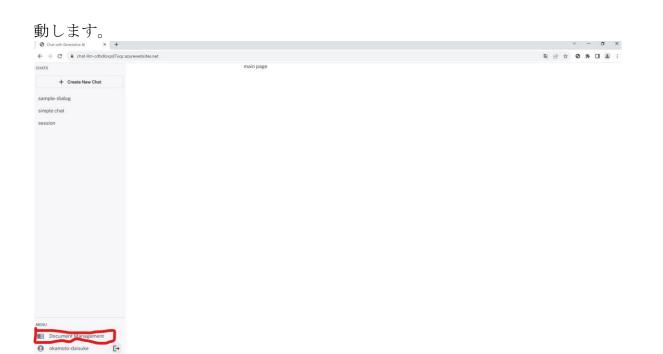
- テキスト文書 (.txt)
- PDF ファイル (.pdf)
- Microsoft Word 文書 (.docx)
- Microsoft PowerPoint プレゼンテーション (.pptx)
- HTML (.html)
- Markdown ファイル (.md)
- CSV ファイル (.csv)
- Microsoft Excel 文書 (.xlsx)

CSV ファイルと Excel 文書は、特定のフォーマットのものに限られます。 具体的には、一列目に各カラムのタイトル、2列目以降にデータが含まれているようなファイルで、それ以外の形式のファイルについては動作を保証しません。 ただし本機能としてはテキストの分割を手動で行い、情報を抽出することが目的であり、それ以外の用途で利用することはできません。

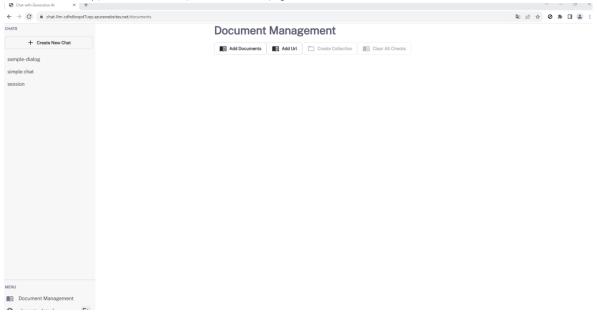
ドキュメント登録



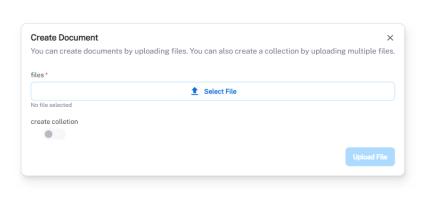
1. 左下の 'Document Management'をクリックし、ドキュメント管理ページに移



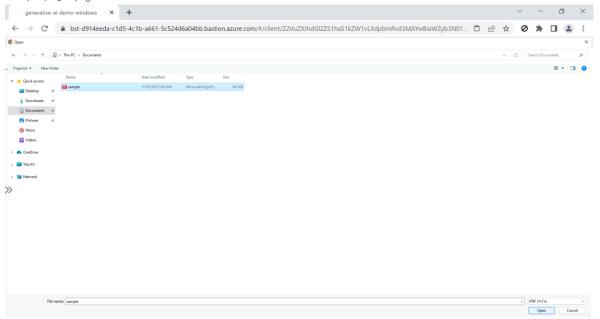
2. Document 管理ページが表示されます。



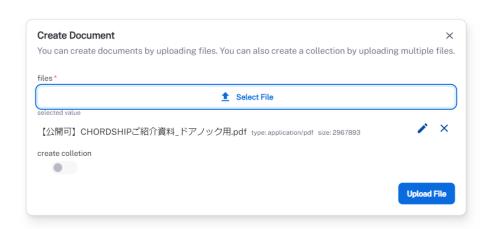
3. 'Add Documents'をクリックするとアップロード用のウィンドウが表示されます。



4. ファイルアップローダーを使用して、登録したいドキュメントをアップロードします。

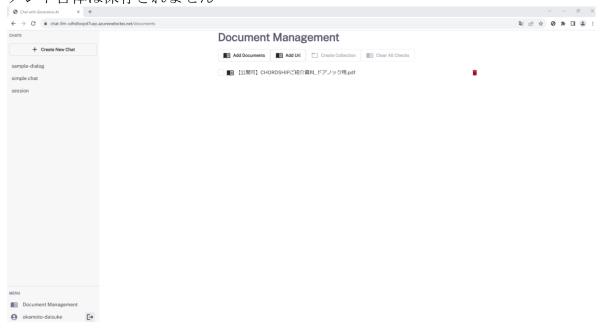


5. 選択したドキュメントが表示されます。

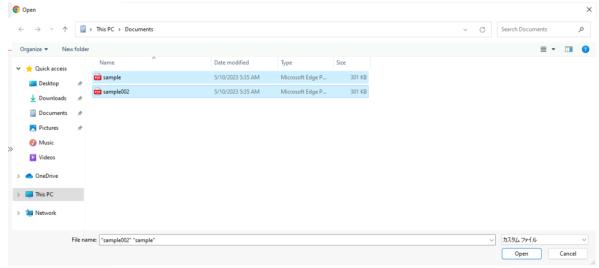


必要に応じてドキュメント名を変更することが出来ます。 (設定しない場合、元のファイル名がそのまま使われます)

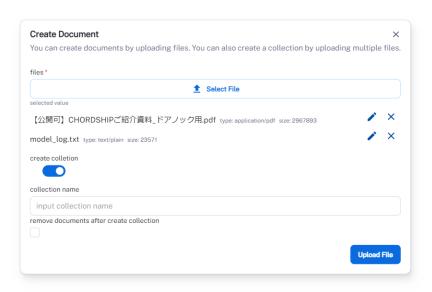
6. 'Upload File' ボタンをクリックします。これによりドキュメントが登録されます。 ※登録時に Azure OpenAI Service の Embedding API を使用してドキュメントがベクトル形式に変換された状態で登録されます。ドキュメント自体は保存されません



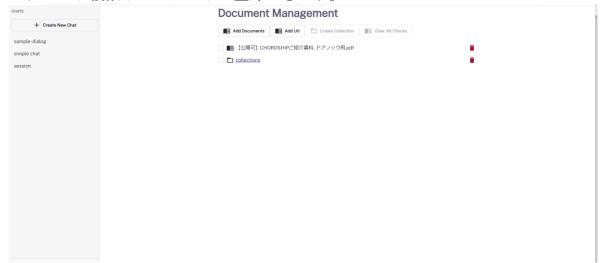
7. ファイルアップロード時に複数のファイルを同時に登録することが出来ま す。アップローダーで複数のファイルを選択してください。



- 8. 'create collection'にチェックを入れることで、アップロードと同時にコレクションにすることが出来ます。
 - コレクションに名前を付けます。(個別のドキュメントに名前を付けることはできません)
 - 'remove documents after create collection'にチェックを入れることでコレクションのみを作成することが出来ます(コレクションの作成後、元のファイルを削除します)



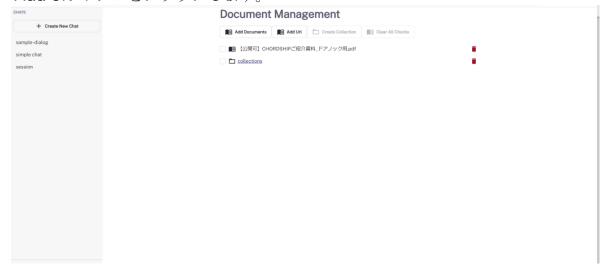
9. これにより複数ドキュメントが登録できます。



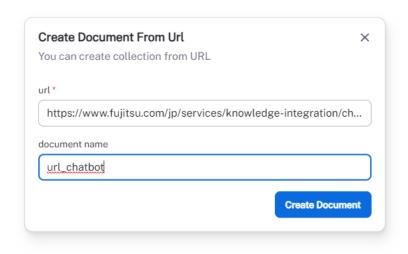
URL からのドキュメントの作成

URL を指定して、その内容からドキュメントを作成できます。

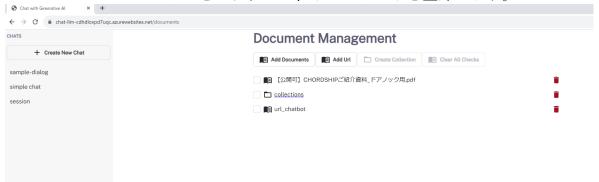
1. 'Add Url' ボタンをクリックします。



2. URL を指定します。必要に応じてドキュメントに名前を付けます。(指定しない場合、URL がそのまま名前となります)



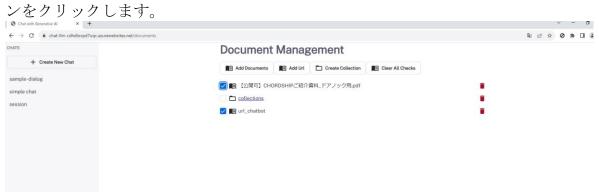
3. 'Create Document' ボタンをクリックし、ドキュメントを登録します。



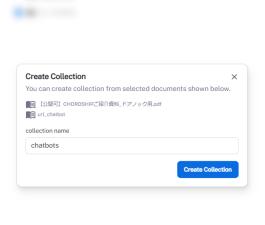
登録済みのドキュメントからのコレクション作成

登録済みのドキュメントを選択して、コレクションを作成できます。

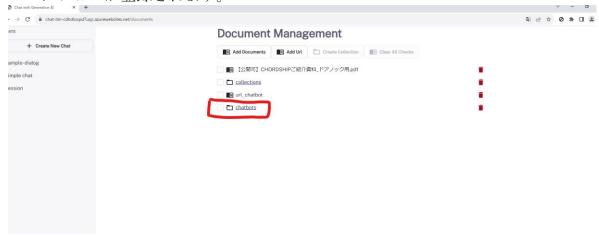
1. 表示されているドキュメントの一覧のうち、コレクションに登録したいドキュメントのチェックボックスにチェックを入れ、'Create Collection'ボタ



登録ウィンドウが表示されます。必要に応じてコレクションに名前を付け 2. ます。'Create Collection'ボタンをクリックし、登録を行います。



3. コレクションが登録されます。



ドキュメントの一覧、削除

登録したドキュメントはゴミ箱アイコンをクリックことで削除できます



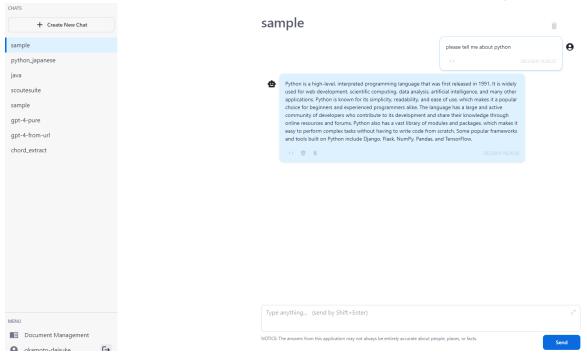
コレクションの右のリンクボタンを押すと、コレクションに含まれるドキュメントを確認できます。



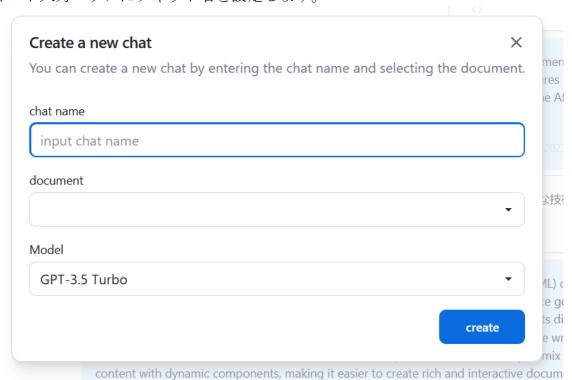
ドキュメントを削除した場合でも、それを含むコレクションからそのドキュメントの内容は失われません。 ドキュメントやコレクションを削除すると、それらを参照しているチャットは新しい会話を続けられなくなります。 ただし、会話の履歴は引き続き参照できます。

4.新しいチャットの作成

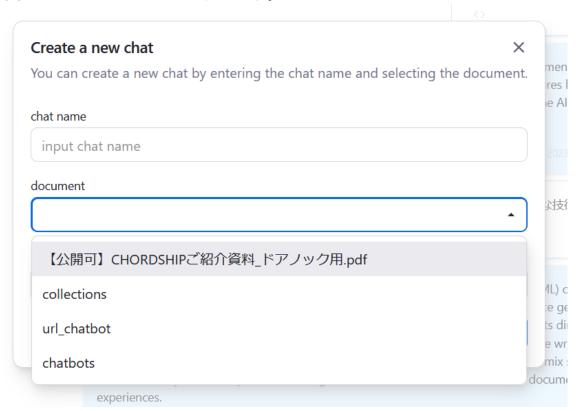
1. サイドバーの上部にある 'Create New Chat' ボタンをクリックします。



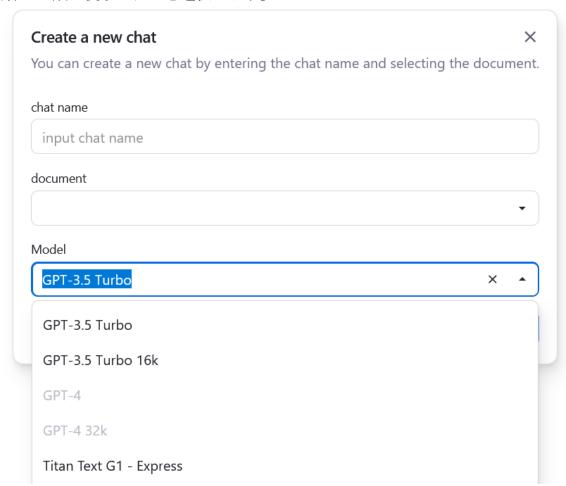
2. テキスト入力エリアにチャット名を設定します。



3. ドキュメントを参照したチャットを行う場合、ドロップダウンメニューから対象となるドキュメントを選択します。

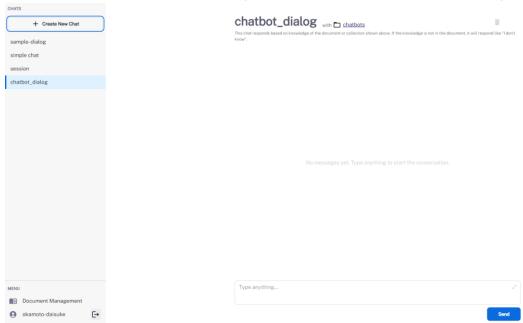


4. 回答の生成に使うモデルを選択します。



グレーアウトされたモデルは利用することができません。

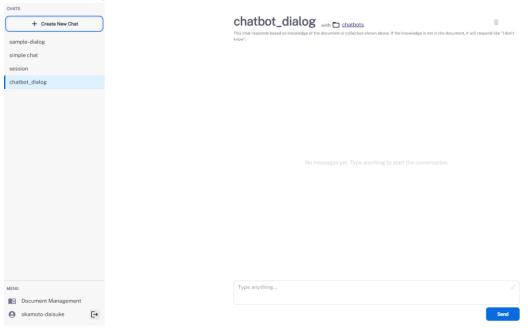
5. 'Create' ボタンをクリックします。新しいチャットの画面が開きます。



ドキュメントを参照しているチャットの場合、チャット名の下にそのドキュメント名が表示されます。 コレクションの場合は、コレクションの詳細画面へのリンクボタンが表示されます。

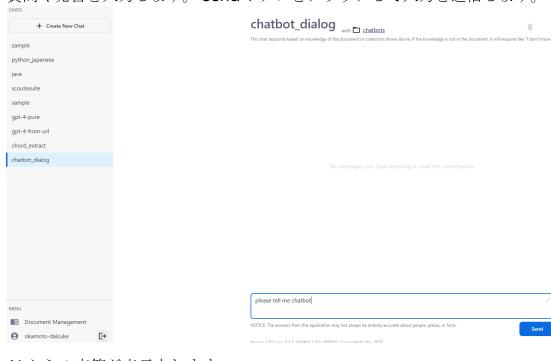
5.チャット

1. サイドバーのチャットタブをクリックします。これによりチャットページ に移動します。

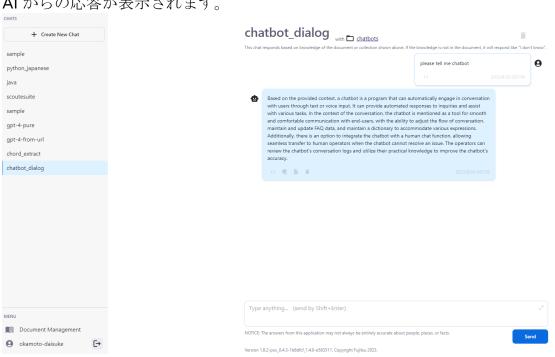


2. チャットに既に会話がある場合は、チャット履歴が読み込まれます。ない場合は、新しい会話を開始できます。

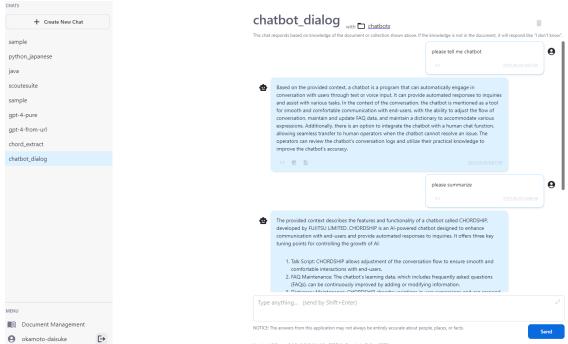
3. AI とチャットするために、チャットページの下部のテキスト入力エリアに 質問や発言を入力します。'Send'ボタンをクリックして入力を送信します。



4. AI からの応答が表示されます。



5. AI が応答した後も、チャットを続けることができます。



- **6.** 会話ウィンドウの左下にあるアイコンクリックで以下のことが実施できます。
 - Based on the provided context, a chatbot is a program that can automatically engage in conversation with users through text or voice input. It can provide automated responses to inquiries and assist with various tasks. In the context of the conversation, the chatbot is mentioned as a tool for smooth and comfortable communication with end-users, with the ability to adjust the flow of conversation, maintain and update FAQ data, and maintain a dictionary to accommodate various expressions. Additionally, there is an option to integrate the chatbot with a human chat function, allowing seamless transfer to human operators when the chatbot cannot resolve an issue. The operators can review the chatbot's conversation logs and utilize their practical knowledge to improve the chatbot's accuracy.

各アイコンの意味合いは左から以下のようになります

- <>マーク:テキスト⇔マークダウンの切り替えを行えます
- 盾マーク:幻覚検出機能を呼び出します

ドキュメントマーク: AI の回答を表示するのに根拠とした登録ドキュメントの参照箇所を表示します



- ゴミ箱ボタン:最後の会話を削除します
- 7. チャット自体を削除したい場合は、チャットのタイトルの右にあるゴミ箱 ボタンで削除することが出来ます。



8. チャットの設定を歯車アイコンの chat setting から変更することが出来ます。



No messages yet. Type anything to start the conversation



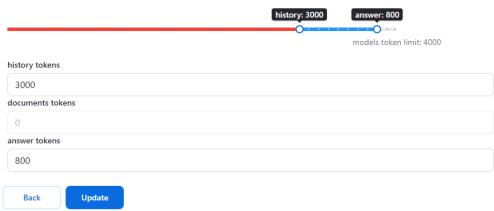
8.1. Tokens Budget

AIモデルが利用するトークンを調整することが出来ます。

test setting

Tokens Budget

You can allocate the token budget to chat history, documents (if any), and tokens that the AI will generate, all within the maximum limits of the model. The remainder can be used for your own statements.



設定可能な項目は以下のようになります。

- history tokens: チャット内の会話履歴の参照に利用する最大トークン
- documents tokens: 参照するドキュメントの最大トークン
- answer tokens: 回答に利用される最大トークン

いずれも、利用されうる最大トークン数を制限するものになります。 例えば answer tokens に 1000 を設定しても 1000 トークン数の出力を保証するものでは ありません。

6. 幻覚検出機能

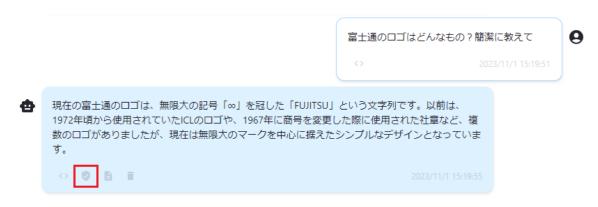
幻覚検出機能を使うと、対話 AI が回答した文章中の「**幻覚**(データに基づかないもっともらしい誤り)」を検出できます。

- 幻覚が検出された場合、幻覚の強さや量に基づいて幻覚スコアと呼ばれる 指標が算出されます。
- ・ 幻覚スコアが 100 に近いほど回答には嘘が含まれている可能性が高いため、 (幻覚検出機能自体の不備・精度不足もありえますが) 信頼できる情報源 を用いた事実確認の要否判断にご活用ください。

ご利用方法

1. ダイアログを開く

検出を行いたい回答枠の下部にある**盾マークのボタン**を押下しますと、幻覚を検 出するためのダイアログが開きます。



(※回答中の「1972 年頃から使用されていたICL のロゴ」は幻覚です。)

2. 機能を選ぶ

タブに表示された3種類の幻覚検出機能の中から、ご利用になりたい1つを選択してください(下図の赤枠内)。

• 各幻覚検出機能は、一度に検出を試みる粒度や方式、使い分け場面が各々 異なります。

「回答全体を多数決」機能:

- 粒度・方式) 多数決に基づいて回答文章全体から幻覚の検出を試み、幻覚 スコアと多数決の結果を表示します。
 - 多数派の意見と無関係か矛盾している回答文章ほど、幻覚スコアが 高くなります。
 - 幻覚が検出された場合、幻覚を除去・緩和した回答文章を表示します。
- **使い分け)** 最も速い一方で検出の粒度は粗いため、以下の場面に適しています。
 - 「大きな幻覚があるか素早く確認したい」
 - 「複数の文にわたる幻覚を見つけたい」

「キーフレーズに注目して多数決」機能:

- ・ **粒度・方式**) 多数決に基づいて回答文章の各文中のキーフレーズ(人物名・年代など)から幻覚の検出を試み、文ごとに幻覚スコアと多数決の結果を表示します。
 - 多数派の意見と無関係か矛盾している回答文ほど、幻覚スコアが高くなります。
 - 幻覚が検出された場合、幻覚を除去・緩和した各文を表示します。
 - タブには、各文の幻覚スコアの平均値が表示されます。
- **使い分け)** 最も遅い一方で検出の粒度は細かいため、以下の場面に適しています。
 - 「細かな幻覚があるか一文一文を確認したい」
 - 「人物名・年代などキーフレーズの幻覚を見つけたい」

「回答と参照文書の対応をチェック」機能:

- 本機能は、ドキュメントを参照するチャット内でのみ利用可能です。
- **粒度・方式)**回答文章とドキュメントとの対応関係を文単位で照合し、文 ごとに幻覚スコアと対応関係の有無を表示します。
 - ドキュメントと無関係か矛盾している回答文ほど、幻覚スコアが高くなります。
 - なお、幻覚の除去・緩和は未対応です。
 - タブには、各文の幻覚スコアの平均値が表示されます。

- **使い分け**) ドキュメントと直接照合するため、以下の場面に適しています。
 - 「ドキュメントに言及が無い回答文があるか確認したい」
 - 「回答文と対応するドキュメント中の文を見つけたい」



3. オプションを選ぶ

幻覚検出機能のオプションを、ダイアログ下部にあるセレクトボックスから1つ 選んでください(下図の赤枠内)。

• このオプションでは、幻覚検出のために回答と比較する **参照(references)** の種類を設定します。

「Use regenerated answers as references」オプション:

- 対話 AI に再度同じ質問をして得た別の回答文章を、比較する参照とします。
- 「回答全体を多数決」機能を利用する時のみ選択できます。

「Use document chunks as references」オプション:

- 対話 AI が回答した際に参照したドキュメントのチャンク(断片)を、比較 する参照とします。
- ドキュメントを参照するチャット内で、「キーフレーズに注目して多数決」 「回答と参照文書の対応をチェック」機能を利用する時のみ選択できます。

「No references」オプション:

- 参照を用いずに、対話 AI が保有する知識だけと比較します。
- 「キーフレーズに注目して多数決」の機能を利用する時のみ選択できます。



4. 幻覚検出を開始する

Check ボタンを押下しますと、選択した幻覚検出機能が実行され、数十秒後に幻覚検出結果が表示されます(下図の赤枠内)。

- ただし、具体的な処理時間は、選択した機能や回答の長さによって変わります。
- ・ 以前の幻覚検出結果を上書きする場合は、代わりに Recheck ボタンが表示 されます。



検出結果の読み方

幻覚スコア

タブや各文頭のチップに表示された数字が、回答文章全体または各文に付けられた**幻覚スコア**です。

- 0~100 の値をとり、100 に近いほど含まれている幻覚の強さや量が大きいことを表しています。
- 幻覚スコアに応じてタブやチップの色が変わります。青色は 0、灰色は 50、 赤色は 100 に対応します。
- ただし、挨拶・謝罪などの幻覚検出できない回答の場合、幻覚スコアは"--"のままタブやチップを黒色で表示します。

「回答全体を多数決」機能

EXPLANATION は、幻覚スコアの算出理由です。多数決の結果が表示されます。 **MITIGATED ANSWER** は、幻覚を除去・緩和した後の回答文章です。

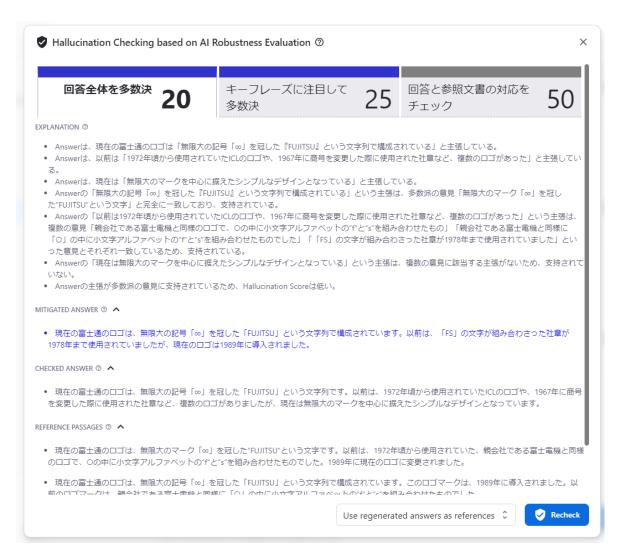
- 基本的に、多数派で合意した意見に修正されます。
- 外部のデータベースやインターネットに問い合わせたものではないため、 修正後の回答文章が一般に認められた事実である保証はありません。

CHECKED ANSWER は、幻覚検出対象の回答文章です。チャット画面に表示されていた回答が再掲されます。

• 表示領域が畳まれている場合、末尾の **v** ボタンを押下してください。

REFERENCE PASSAGES は、多数決する際に参照した別の回答文章です。対話 AI に同じ文脈で3回再回答させた結果が表示されます。

• 表示領域が畳まれている場合、末尾の **v** ボタンを押下してください。



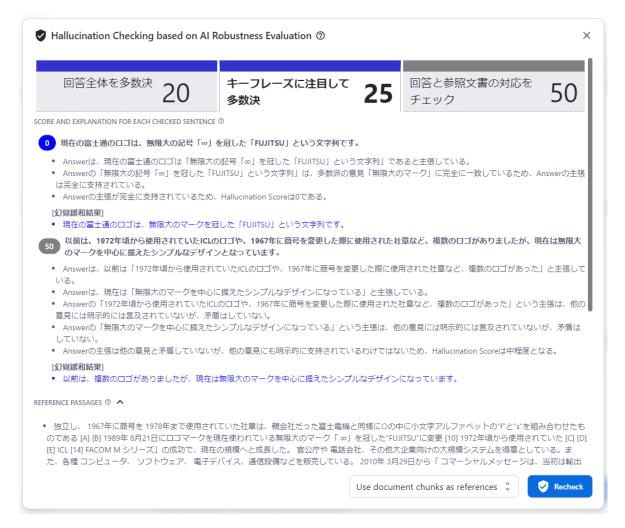
「キーフレーズに注目して多数決」機能

SCORE AND EXPLANATION FOR EACH CHECKED SENTENCE は、文ごとの幻覚スコアと算出理由です。

- 算出理由として、キーフレーズごとの多数決の結果が表示されます。
- 幻覚が検出された場合、幻覚を除去・緩和した後の文を、「幻覚緩和結果」 にそれぞれ表示します。
 - 基本的に、多数派で合意した意見に修正されます。
 - 外部のデータベースやインターネットに問い合わせたものではない ため、修正後の回答文章が一般に認められた事実である保証はあり ません。

REFERENCE PASSAGES は、キーフレーズを多数決する際に参照したドキュメントのチャンク(断片)です。

- 「Use document chunks as references」オプションを選択した時のみ表示されます。
- 表示領域が畳まれている場合、末尾のマボタンを押下してください。

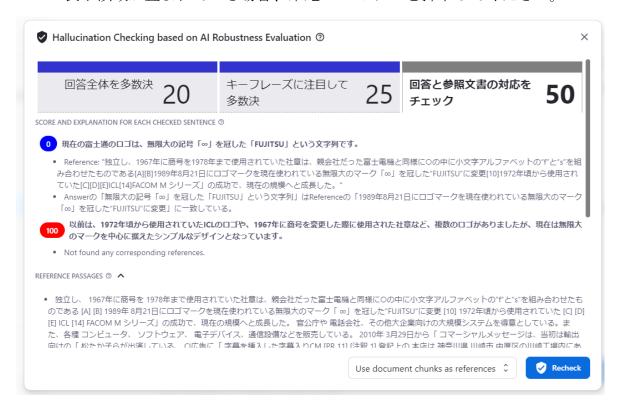


「回答と参照文書の対応をチェック」機能

SCORE AND EXPLANATION FOR EACH CHECKED SENTENCE は、文ごとの幻覚スコアと算出理由です。

• 算出理由として、対応関係があったドキュメントの文とその関係性が表示 されます。 **REFERENCE PASSAGES** は、対応関係を照合する際に参照したドキュメントのチャンク(断片)です。

表示領域が畳まれている場合、末尾の V ボタンを押下してください。



7. フィッシング URL 検出機能

フィッシング URL 検出機能を使うと、対話 AI との対話文章中にある URL が、情報を盗むために悪意ある攻撃者が作成したサイトの URL(フィッシング URL)かどうかを判定して提示できます。 フィッシング URL が検出された場合、該当する URL の横にエクスクラメーションマーク(!)が表示され、URL をクリックできなくなります。本機能により、対話 AI をより安心して扱うことが可能になります。

ご利用方法

基本機能

自動的に動作します。対話中に URL が存在する場合、それぞれの URL に対してフィッシング URL かどうかを判定します。



アメリカの観光名所のいくつかとそのURLを教えますね: 1. グランドキャニオン国立公園(アリソナ州) - www.nps.gov/grca/index.htm 2. ヨセミテ国立公園(カリフォルニア州) - www.nps.gov/yose/index.htm 3. 自由の女神像(ニューヨーク) - www.nps.gov/stii/index.htm 4. ハワイ火山国立公園(ハワイ州) - www.nps.gov/havo/index.htm 5. ワシントン記念塔(ワシントンD.C.) - www.nps.gov/wamo/index.htm 6. ラスベガスのストリップ(ネバダ州) - www.niagarafallsstatepark.com 7. ニアガラの滝(ニューヨーク州/オンタリオ州) - www.niagarafallsstatepark.com 8. ハリウッドウォークオブフェーム(カリフォルニア州) - www.walkoffame.com 9. ディズニーランド(カリフォルニア州) - www.disneyland.disney.go.com 10. マウントラッシュモア国立記念公園(サウスダコタ州) - www.nps.gov/moru/index.htm これらはアメリカの人気のある観光地の一部です。アメリカにはさまざまな魅力的な場所がありますので、ぜひ訪れてみてください!

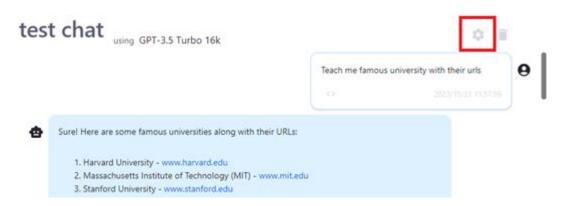
- 2. URL に危険性がない場合、そのままリンクとして表示されます。
 - 1. グランドキャニオン国立公園(アリゾナ州) www.nps.gov/grca/index.htm
 - 2. ヨセミテ国立公園(カリフォルニア州) www.nps.gov/yose/index.htm
 - 3. 自由の女神像(ニューヨーク) www.nps.gov/stli/index.htm
- 3. フィッシング URL の可能性が高い場合、URL の横にエクスクラメーション マークが表示され、クリックできなくなります。

https://www.eki-net.faegardens.com 🛕

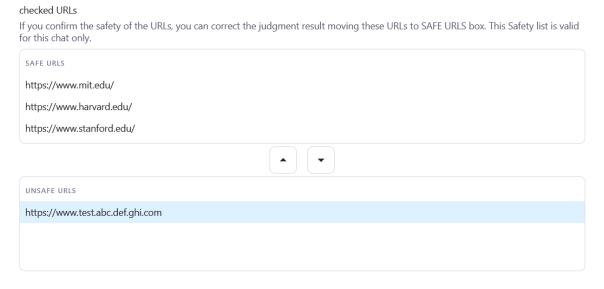
URL 判定結果の修正機能

利用者の確認などにより、フィッシング URL の可能性が高い URL の安全性が確認できた場合、判定結果を修正することができます。

1. ウィンドウ右上のマーク「chat setting」をクリックしてください。



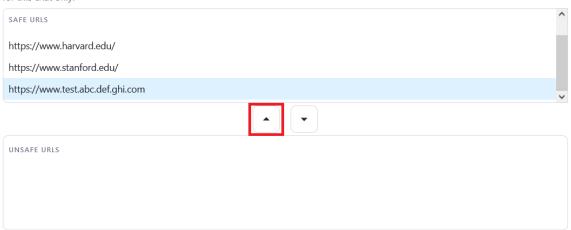
2. 「checked URLs」の項目で、安全が確認できた URL を「UNSAFE URLS」から選択します。



3. 上向きボタンで「SAFE URLS」に移動させます。

checked URLs

If you confirm the safety of the URLs, you can correct the judgment result moving these URLs to SAFE URLS box. This Safety list is valid for this chat only.



4. 「Update」ボタンで安全な URL として登録され、クリックが可能になります。

フィッシングURL 検出機能のオン/オフの切り替え

「chat setting」内でフィッシング URL 検出機能のオンオフを切り替えられます。 「enable phishing URL check」の項目でトグルボタンをクリックして切り替えてください。

オンの状態

enable phishing URL check

You can change the state of the Phishing URL feature enable or disable.



オフの状態

enable phishing URL check

You can change the state of the Phishing URL feature enable or disable.



8.トラブルシューティングと FAQ

問題が発生し、解決しない場合は、当社担当者にお問い合わせください。

9.アプリケーションの終了

アプリケーションの使用が終わったら、ウェブブラウザやアプリケーションを実行しているタブを閉じることで終了できます。