

Generative AI Cloud (SaaS)

サービス仕様書

12 版

2026 年 3 月 27 日 日本電気株式会社

改版履歴

版数	日付	改版内容
1	2024 年 12 月 26 日	初版
2	2025 年 3 月 31 日	<ul style="list-style-type: none">・新提供機能：ハルシネーション対策機能の諸元を追加・会話履歴の保存期間の諸元を追加・サポート対象外の内容を追加で明文化・Agentic AI for Search を提供機能に追加
3	2025 年 4 月 25 日	<ul style="list-style-type: none">・新提供機能：図表文脈理解機能の諸元を追加
4	2025 年 6 月 6 日	<ul style="list-style-type: none">・cotomi モデルのバージョンアップに伴う旧バージョンの提供を規定
5	2025 年 8 月 8 日	<ul style="list-style-type: none">・Azure OpenAI 提供モデルの終了時期の把握と対応について規定
6	2025 年 10 月 28 日	<ul style="list-style-type: none">・追加ユーザ 1000、多要素認証オプションをサービスメニューに追加・提供機能に AI ガードレールを追加・機能および API の諸元に AI ガードレール、多要素認証を追加・ユーザアカウントの種類と役割に追加された役割を追加
7	2025 年 12 月 4 日	<ul style="list-style-type: none">・秘匿問い合わせ API プレビュー、画像入力、Claude モデルを追加・「8.提供機能」にてチャット UI を利用する場合、利用者ごとにユニークなメールアドレスが必要な旨、追記。
8	2025 年 12 月 18 日	<ul style="list-style-type: none">・AI ガードレール（プレビュー）、および秘匿問い合わせ（プレビュー）の提供終了に伴い、関連記述を削除。
9	2025 年 12 月 26 日	<ul style="list-style-type: none">・AI ガードレールの正式対応の反映、また AI ガードレール利用における注意事項を記載するため、「7.注意事項」を追加・「5 サービスメニュー」の基本メニューの発行コール数を cotomi v3 に合わせて変更
10	2026 年 2 月 13 日	<ul style="list-style-type: none">・ポリシー違反についての注意事項を「7.注意事項」に追加
11	2026 年 2 月 28 日	<ul style="list-style-type: none">・機微情報の秘匿化 API（プレビュー版）を追加・サービス提供時間の通知手段として管理ポータルへのお知らせ表示を追加
12	2026/3/27	<ul style="list-style-type: none">・「10 機能および API の諸元」の「インデックス管理機能」に関連チャック検索 API の諸元を追加・「10 機能および API の諸元」、「7. 注意事項」に図表文脈理解機能の注意事項を追加

目次

Generative AI Cloud (SaaS)	1
サービス仕様書	1
1. 本文書について	4
1 本文書の目的	4
2 本文書の改版	4
3 用語集	5
2. サービス体系	7
4 サービス概要	7
5 サービスメニュー	8
6 契約単位の考え方	10
7 従量課金の考え方	10
3. サービス内容	11
8 提供機能	11
9 提供モデルおよび諸元	12
10 機能および API の諸元	13
11 ユーザアカウントの種類と役割	15
12 サービス提供時間	15
4. サポート	16
13 動作環境	16
14 保守サポート	16
15 本サービスのアップデート	16
16 本サービスのバックアップ	17
17 本サービスの責任範囲	17
5. 利用申し込み・変更・終了と利用料金について	18
18 利用契約の申し込みと利用料金発生のタイミング	18
19 利用契約の変更と変更完了のタイミング、利用契約の終了について	18
20 従量課金について	18
6. 適用される各社のプライバシーポリシー	19
7. 注意事項	19

1. 本文書について

本文書は、秘密保持対象ドキュメントとして、サービス事業者の許可なくコピーおよびその配布、ホームページへの掲載などの第三者への開示を禁じます。

1 本文書の目的

本文書は、契約者 に サービス事業者 が提供する「Generative AI Cloud (SaaS)」 (以下「本サービス」といいます。) の仕様を記載した文書です。

2 本文書の改版

本文書の見直しは、本サービスの提供内容の変更に伴い実施します。
契約者は、その時点での最新のサービス仕様書を参照し、仕様を確認ください。

3 用語集

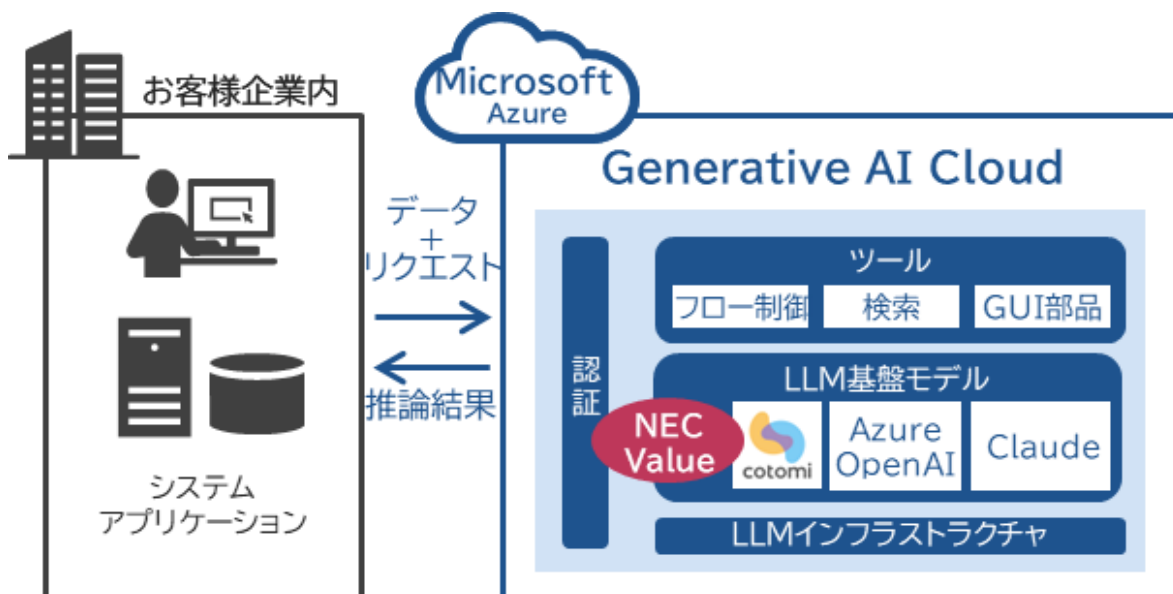
用語（別称、略語）	説明
Generative AI Cloud (SaaS) (本サービス)	Generative AI Cloud (SaaS)利用約款に基づき当社が契約者に提供する本書で定めるサービスを意味します。
サービス提供者	自身のサービスを提供する目的のために、当社との間で、利用約款に基づく利用契約を締結した法人その他の団体をいいます。
サービス利用者	サービス提供者が提供するサービスを利用する個人、法人およびそのたの団体をいいます。
契約者	サービス提供者、および、当社との間で利用約款に基づく利用契約を締結した法人その他の団体をいいます。
利用者	契約者における個人をいいます。
利用契約	利用約款に基づき当社と契約者との間に締結される本サービスの提供に関する契約をいいます。
ユーザID	本サービスの利用にあたり当社から契約者に付与され、契約者とその他の者とを識別するために用いられる符号をいいます。
パスワード	本サービスの利用にあたりユーザが設定するもので、ユーザIDと組み合わせて、契約者とその他の者とを識別するために用いられる符号をいいます。
APIキー	本サービスの利用にあたり当社から契約者に付与され、契約者がインターネットを経由して本サービスに接続するために用いられる符号をいいます。
契約ポータル	NEC Cloud Services ポータル内にある Web サイトです。 本サービスの契約申込み、契約変更申込みおよび契約状況の閲覧等を行うための機能を提供します。
運用ポータル	本サービスを利用するために必要な設定情報の登録機能を有する Web サイトです。
営業日	日本国における土日祝祭日および当社所定の休業日を除く日をいい、営業時間は 9:00~17:00 (JST) とします。
契約者データ	本サービスの利用に伴い、本サービス用設備に契約者が保管するデータをいい、プロンプト、出力結果その他契約者の秘密情報を含みます。
LLM	本サービスの一部として当社が契約者に提供する、大規模なデータセットとディープラーニング技術を含む機械学習技術を用いて構築された学習済みの言語モデル(契約者と当社との間で別途LLMのファインチューニング等に係る契約が締結された場合、当該契約に基づき当社がチューニングしたモデルを含みます)をいいます。

プロンプト	利用者が LLM に入力する指示・条件・参照情報等の文字列です。
プロンプトファイル	本サービスの一部として当社が契約者に提供するファイルをいいます。
cotomi	NEC が提供する LLM の商標です。
ファインチューニング	あらかじめ存在するモデルに利用者が追加学習させて、新しいモデルを生成することをいいます。
アクティブ化	登録したモデルはアクティブ化を行うことで API から指定して利用することができます。アクティブ化できる数の条件は本サービス Generative AI Cloud (SaaS) BYOM 実行環境費の契約数で決まります。
度数	度数は従量課金を計算する際の数量になります。単金×度数が従量課金の請求額として請求されます。 度数は利用したモデルとトークン数毎に算出式を持ち計算されます。具体的な計算方法はマニュアル（利用可能 LLM）に記載しています。

2. サービス体系

4 サービス概要

本サービスは、当社が運営する LLM に API を通じてリクエストを投げるだけで推論結果が得られるクラウドサービスです。本サービスを利用することで、LLM を動作させるインフラを新たに準備・開発する必要なく、契約者の有するサービスやシステムに LLM 機能を組み込むことができます。



5 サービスメニュー

本サービスのメニュー体系は以下の通りです。各メニューの利用料金については、価格表を参照ください。

本サービス利用にあたっては、基本メニューのいずれか1つの契約が必要となります。

メニュー		説明・諸元	
基本メニュー	Generative AI Cloud (SaaS) スターター	最大コール数 (※1)	12 コール/分
		登録ユーザ数	10 ユーザ
	Generative AI Cloud (SaaS) ベーシック	最大コール数 (※1)	40 コール/分
		登録ユーザ数	80 ユーザ
オプションメニュー	Generative AI Cloud (SaaS) BYOM 実行環境費	サービス利用者専用のモデル利用環境。 基本メニューの購入ならびに、実行するモデル環境数分の購入が必要。	
	情報検索(RAG)	以下の機能を利用する際に購入が必要 ・ 情報検索の機能 ・ Agentic AI for Search (※3) ・ 図表文脈理解機能	
	情報検索(RAG) 拡張ストレージ	インデックスに登録できる文書のディスク容量	
	追加ユーザ 1 0 0 0	ベーシック契約において、ユーザが 80 を超える場合に必要。本オプション 1 契約で 1,000 ユーザ追加可能。	
	多要素認証	多要素認証によるログインを希望される場合に必要。 本オプションは新規契約時に選択必要。利用開始後に追加することはできません。	
従量課金メニュー	トークン使用量	トークンの使用量 (事後精算)	

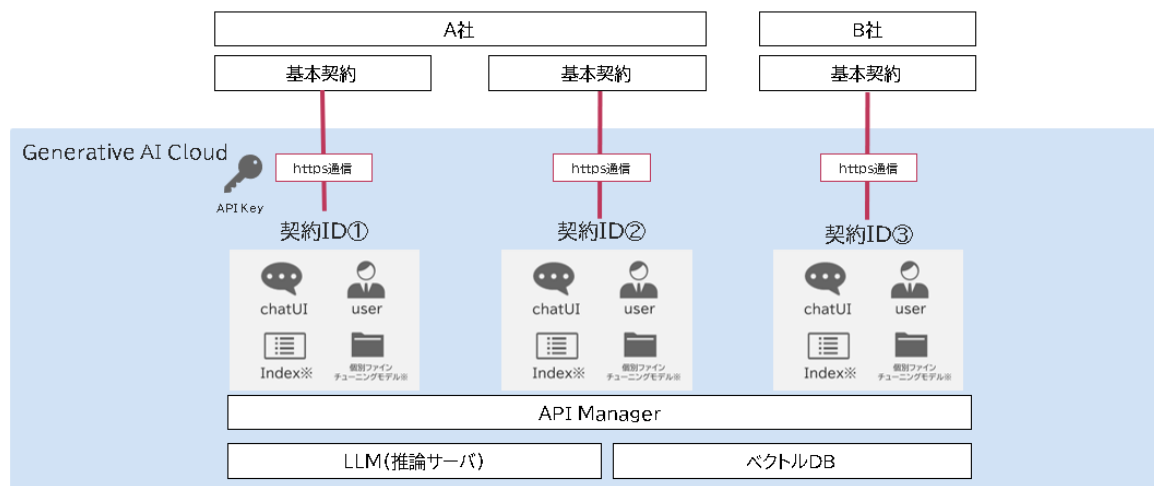
※1 契約時に発行された API キーおよび Generative AI 管理ポータルから払い出したユーザを用いて、未完了のリクエストを同時に発行できる上限です。本諸元は cotomi v3 の場合の諸元です。他社モデルの諸元については当社にお問合せください。

※2 ファインチューニングの実施は、別途定める約款・仕様書に基づく契約が必要となります。利用にあたっては、当社にご相談ください。

※3 Agentic AI for Search の利用には情報検索(RAG)の他に基本メニューとしてベーシックの契約が必要となります。

6 契約単位の考え方

本サービスにおいて、当社は基本契約ごとに契約 ID を割り当て、契約管理者アカウントを設定します。また、基本契約ごとに API キーが払い出されます。



7 従量課金の考え方

本サービスが提供する API から LLM のモデルが実行される場合にトークン数に応じた課金します。カウントされる対象は リクエストで受け付けるトークン数とレスポンスで生成したトークン数です。オプションメニューの BYOM 実行環境で実行するモデルの使用につきましては、オプション費用に含まれておりますので、従量課金の対象外です。

トークン数は 1 日 1 回以上で定期的集計されます。利用したトークン数については、管理ポータルから確認できます。具体的な確認方法についてはマニュアルを参照ください。

3. サービス内容

8 提供機能

本サービスが提供する機能（以下、提供機能といいます）は、下記の通りです。

機能	Generative AI Cloud (SaaS) スターター	Generative AI Cloud (SaaS) ベーシック
LLM モデル(※1)	・ cotomi ・ Azure OpenAI 提供モデル ・ Claude	・ cotomi ・ Azure OpenAI 提供モデル ・ Claude
API	○	○
チャット UI	×	○(※2)
管理機能	△（グループ機能は利用不可）	○

(※1) 提供しているモデルの最新バージョンは、当社までお問合せください。

(※2) チャット UI を利用するには利用者ごとにユニークなメールアドレスが必要です。

機能	説明
Agentic AI for Search	情報検索(RAG)およびベーシックを契約することで利用できます
図表文脈理解機能	情報検索(RAG)を契約することで利用できます
AI ガードレール	API、およびチャット UI で利用できます
機微情報の秘匿化 (プレビュー)	API で利用できます
画像入力	画像ファイル (jpeg/png/webp/gif) をもとにした API およびチャットが可能です (利用するモデルが画像入力に対応している必要があります)。

9 提供モデルおよび諸元

本サービスで提供する LLM モデルと各モデルの諸元につきましては、ユーザガイドを参照ください。
なお、cotomi モデルについて、新バージョンリリース後、半年後に旧バージョンの提供を終了予定です。
Azure OpenAI 提供モデルは Microsoft 社の提供方針に則ります。そのため利用中のモデルの提供終了
時期や後継モデル等は利用者にて Microsoft 社のサイトをご確認ください。

<https://learn.microsoft.com/ja-jp/azure/ai-foundry/openai/concepts/model-retirements>

注：本 URL は変更になる可能性があります。

また、Claude のモデルについては当社にお問い合わせください。

10 機能および API の諸元

本サービスが提供する機能および API に関する諸元です

機能	諸元
拡張対話（ファイル添付）機能	添付可能なファイル容量の上限は 30MB とします。
インデックス管理機能	作成できるインデックス数の上限は 1 万件とします。
インデックス管理機能	1 インデックスあたりの文書登録数の上限は 1 万件とします。
インデックス管理機能	文書登録のファイルサイズは 16MB(ファイル内のテキストサイズは 1MB を上限、ただし xls,xlsx については 500KB を上限) とします。
インデックス管理機能	ユーザガイド記載の非同期の文書登録コマンドについては、契約単位で 1 日当たりの使用量について以下の諸元があります。 1. データサイズ 1 日当たりの投入データの合計サイズ：最大 100MB 2. ファイル数 1 日当たりの投入ファイルの合計数：最大 200 ファイル
インデックス管理機能	関連チャック検索 API における検索数の指定 (topK) について、推奨値：100、最大値：5000 とします。詳細はマニュアルの「ベクトル DB 管理 API チュートリアル」を参照ください。また、非推奨の値を指定して利用する場合は、最大コール数の上限は標準値の 1/5 とします。最大コール数については「5.サービスメニュー」を参照ください。
回答根拠の確認機能（推論なし）	回答、参照文書それぞれの全体の文章の長さは 5 万字、文章内の文章数（文の数）は 250 までを上限とします。
回答根拠の確認機能（推論なし）	回答、参照文書それぞれの 1 文の長さは 512~1024 文字を上限とします（判定上限の文字数は文章によって異なります）
推論を用いた回答根拠の確認機能	回答文章の長さの上限は 1 万字、回答文章、参照文章の長さの合計は 10 万字を上限とします。
会話履歴の保存期間	対話機能全般での会話履歴の保存期間は会話開始時から 180 日(約 6 ヶ月)になります。期間を過ぎた履歴は削除され参照できなくなります。
図表文脈理解機能	図表文脈理解機能に対して同時実行はできません。
図表文脈理解機能	10 分間あたり 10 ページの文書データの実行を上限とします。 ※ 本諸元は 1 ページあたりの図表数が 1 個程度の文書を想定しています。 本諸元を遵守している場合においても本機能の利用状況によってはリクエストが失敗する可能性があります。
図表文脈理解機能	図表文脈理解機能は、入力ファイルの図や表を含めた記載内容をテキストにして出力しますが、記載されている全ての情報が漏れなく正確に出力されるとは限りません。必要に応じて、意図通りの情報が出力されているかを確認してください。
AI ガードレール	API、およびチャット UI で利用できます。 AI ガードレールはリスクの低減を目的とした機能で、すべての不適切なコンテンツを完全に防止できるものではありません。 入力および出力内容について、利用者自身が問題ないかを確認する必要があります。

多要素認証	以下の Authenticator をサポートします。 <ul style="list-style-type: none">• Google Authenticator• Microsoft Authenticator• FreeOTP 本オプションは新規契約時に選択必要。利用開始後に追加することはできません。
画像入力	画像ファイル (jpeg/png/webp/gif) をもとにした API およびチャットが可能です (利用するモデルが画像入力に対応している必要があります)。
機微情報の秘匿化 (プレビュー)	API でのみ利用可能です。 また、機微情報の秘匿化 API は完全な秘匿化を保証するものではありません。機微情報の秘匿化 API の結果は利用者で問題ないか確認が必要です。

11 ユーザアカウントの種類と役割

API 管理ポータルアカウントには複数の種類があります。それぞれの役割を以下の表に示します。

役割（ロール）	内容
組織管理者	管理 WebUI の全ての操作が行える役割
ユーザ管理者	管理 WebUI でユーザの一覧参照、追加、変更、削除が行える役割
グループ管理者	管理 WebUI でグループの一覧参照、追加、変更、削除が行える役割
インデックス管理者	管理 WebUI でインデックスの一覧参照、追加、変更、削除、および、インデックスにユーザ、グループの紐づけが行える役割
ドキュメント管理者	管理 WebUI でインデックスの一覧参照、文書管理、および登録文書一覧参照、文書の追加、削除が行える役割
テンプレート管理者	管理 WebUI でテンプレートの一覧参照、追加、員表、編集、削除、およびテンプレートにグループの紐づけが行える役割
一般ユーザ	Generative AI API サービスを利用できる役割 チャット UI を利用できる役割

12 サービス提供時間

本サービスの機能は、計画停止時を除き、24 時間 365 日利用可能です。

計画停止は NEC Cloud Services ポータルのお知らせ機能、および管理ポータルのお知らせ表示で通知します。

4. サポート

13 動作環境

Web アプリケーションが動作する Web ブラウザは以下です。

バージョンの詳細はユーザガイドを参照ください。

Microsoft Edge

Google Chrome

14 保守サポート

保守サポートの内容は以下の通りです。

- 対応時間帯
 - ・ 当社の営業日
- 具体的な提供内容
 - ・ 本サービスの利用方法に関する問合せへの回答
 - ・ 本サービスの各種ドキュメント記載事項に関する問合せ・回答
 - ・ 本サービスの不具合に関する調査・回答
- サポート対象外
 - ・ Generative AI Cloud(SaaS)の具体的な設定方法
 - ・ プロンプトの書き方や各モデルの回答についての問い合わせ
 - ・ 推論結果（内容）の正確性ならびに応答速度に関する問合せ・回答

15 本サービスのアップデート

当社は、本サービスの機能強化や不具合修正に伴い、提供機能（管理ポータル画面を含む）のアップデートを行う場合があります。契約者は提供機能（管理ポータル画面を含む）の最新バージョンのみを利用できます。

16 本サービスのバックアップ

当社が本サービス用設備障害時の環境復旧のために実施するバックアップは以下の通りです。

対象	データ保存期間
管理ポータル画面で 設定する情報や履歴情報	7日間

17 本サービスの責任範囲

Generative AI Cloud (SaaS) 利用約款の定めに加え、本サービスにおける 当社及び契約者の責任範囲を次に示します。

対応者	責任範囲
当社	<ul style="list-style-type: none">・ 契約者へ本サービスを提供するために必要な本サービス用設備およびその他の環境の構築・ 本サービスの運用時の本サービス用設備およびその他の環境のリソース拡張・ 本サービスの不具合の問合せ対応・ 本サービスの不具合の調査、復旧・ 本サービス用設備障害時の環境復旧のためのバックアップおよび復元作業・ 第三者による攻撃を防ぐためのセキュリティ対策
契約者	<ul style="list-style-type: none">・ 契約者設備その他 本サービスの利用に必要な環境の準備・ 本サービスを利用するための運用ポータル上での設定・ 利用者のシステムアプリケーションへの本サービス組込みの設定・ 利用者固有のデータ（定義情報等）、誤操作によるデータ削除時の復旧用データのバックアップおよび復元作業・ 第三者等による 不正利用を防止するための アカウントおよび受信用メールアドレス情報等の管理・ サービス利用・変更・削除申請・ Azure OpenAI 提供モデルの終了時期の把握とそれに応じた必要な対応・ 機微情報の秘匿化 API の呼び出しとその結果の利用（※1） <p>（※1）プレビュー版では秘匿化と復元の単体機能を API で提供しています。秘匿化した結果の LLM への問い合わせや、その LLM からの回答を復元する処理は機微情報の秘匿化 API を利用して利用者で作りこみが必要です。</p> <p>また、機微情報の秘匿化 API は完全な秘匿化を保証するものではありません。機微情報の秘匿化 API の結果は利用者で問題ないか確認が必要です。</p>

5. 利用申し込み・変更・終了と利用料金について

本サービスの契約方法と利用料金の考え方について説明します。

18 利用契約の申し込みと利用料金発生タイミング

契約者は本サービスを利用する場合、契約ポータルからサービスメニューを申し込みます。利用契約成立後、サービス事業者は管理ポータルへの招待メールを送付します。本メールが発信された時点から契約者は本サービスを利用できます。

また、契約ポータルから別途準備完了のメールを送付します。利用料金が発生します。本メールが発信された時点から利用料金が発生します。

19 利用契約の変更と変更完了のタイミング、利用契約の終了について

利用契約を変更する場合、契約者は契約ポータルから変更の申請を行う必要があります。本申請に対して事業者が契約ポータルから承諾を行ったタイミングで変更が成立します。変更の申請から承諾まで最大 3 営業日かかるため、当月内の変更完了を希望する場合は変更 希望月最終営業日の 3 営業日前までに実施する必要があります（※）。

利用契約は、すべての基本契約とオプション契約の解約手続きが終了となる変更が完了した時点で終了します。利用契約の終了の翌月以降は利用料金が発生しません。

（※）例：最終営業日が 31 日（金）の場合、3 営業日前は 28 日（火）です（29 日、30 日は通常営業の場合）。最終営業日が 31 日（月）の場合は、3 営業日前は 26 日（水）です。

20 従量課金について

- ・ 本サービスでは、利用したトークン数から度数に変換され、度数×単価で請求額を決定します。
- ・ 度数に換算する際のレートは、利用するモデル毎にリクエストで受け付けるトークン数とレスポンスで生成したトークン数のそれぞれで異なります。具体的なレートの値はマニュアルでご確認ください。
- ・ 当月分の利用料は翌月の支払いで請求されます。

6. 適用される各社のプライバシーポリシー

本サービスでは以下クラウドサービスを利用しており、本サービス利用にあたり、各社が規定するプライバシーポリシーが適用されます。

クラウドサービス	処理内容	処理に利用されるデータ	所在	プライバシーポリシー
Microsoft Azure	本サービスのシステム利用に必要なデータの保管	本サービスで扱う情報全般	日本	https://azure.microsoft.com/ja-jp/explore/trusted-cloud/privacy

7. 注意事項

- (1) AI ガードレールはリスクの低減を目的とした機能で、すべての不適切なコンテンツを完全に防止できるものではありません。入力および出力内容について、利用者自身が問題ないかを確認する必要があります。
- (2) 本サービスでは利用約款 第 21 条で以下の疑義のある行為を禁止しております。
(禁止事項) 第 2 1 条 契約者は、本サービス等を利用するにあたり、次の各号に規定する事項を行わないと共に、これらに関する疑義等を生じさせず、かつ当社による本サービス等の提供に支障の生じることのないようにするものとします。
 - (2 1) その他法令もしくは公序良俗に違反し、または当社もしくは第三者に不利益を与える行為
 - (2 2) 前各号のいずれかに該当するおそれがあると当社が判断する行為
 - (2 3) その他本サービスの運営を妨げると当社が判断する行為

上記は各モデルに関するプロンプトの内容にも適用されます。第 2 1 条に違反する内容、および疑われる内容の入力・利用は行わないようにしてください。ポリシー違反が検出された場合、当該テナントの利用を停止させていただく場合がありますのでご注意ください。

なお、本サービスが提供する他社クラウドベンダのモデルに関しては、モデル提供先でコンテンツフィルターによるポリシー違反の自動監視機能がある場合に本サービスでは監視を有効にしております。

これらのフィルターにて当該行為の検知が続く場合には当該テナントの利用を停止させて頂く可能性があります。

Azure など提供されるモデルをご利用の際は、必ず各モデルの提供先の利用規約(コンテンツの要件や不正使用の監視)についてもご確認ください。

Microsoft Azure

- ・ コンテンツ要件

<https://learn.microsoft.com/ja-jp/legal/ai-code-of-conduct?context=/azure/ai-foundry/openai/context/context>

- ・ 不正利用監視

<https://learn.microsoft.com/ja-jp/azure/ai-foundry/openai/concepts/abuse-monitoring?view=foundry-classic>

Claude

- ・ 不正使用検出

<https://docs.aws.amazon.com/bedrock/latest/userguide/abuse-detection.html>

- (3) 図表文脈理解機能は、入力ファイルの図や表を含めた記載内容をテキストにして出力しますが、記載されている全ての情報が漏れなく正確に出力されるとは限りません。必要に応じて、意図通りの情報が出力されているかを確認してください。

商標について

本文書に掲載された各社名、各製品名は、各社の商標または登録商標です。