

Generative AI FW V2.2.2 リリースノート

1. はじめに

Generative AI FW V2.2.2のリリースノートです。

1.1. 対象バージョン

Generative AI FWのバージョン情報は以下の通りです。

確認対象	バージョン
チャット画面	build 202512112355 または build 202512080254
管理ポータル	build 202601190519または build 202512110856

バージョンの確認手順は後述の「参考：バージョンの確認方法」を確認してください。

2. リリース内容

2.1. 動作環境

2.1.1. LLMのサポート範囲変更

製品がサポートするLLMはcotomi v3のみに変更になります。cotomi Fast、cotomi Proは使用できません。

2.1.2. 対象OS、GPUの追加

美琴のHW変更に伴い、動作環境に以下が追加されました。従来の動作環境も引き続きサポートいたします。

対象OS : Red Hat Enterprise Linux 9.6

GPU : NVIDIA RTX PRO 6000 SE

2.1.3 cotomi のアップデート

新しいHWに対応するため、cotomi v3.0.1に対応します。アップデート時はcotomiのアップデートも必要になります。

使用するOSSのアップデートによりキャッシュ機能が追加されています。

2.1.4. OSSの最新化

Generative AI FWで使用するOSSを脆弱性に対応するために最新化しています（2025年12月時点）。

2.2. 新機能

2.2.1. AIガードレールオプションの追加

Generative AI FW AIガードレールオプションが追加されました。LLMを使った対話時に入力および出力内容の安全性チェックを行い、不適切なコンテンツのブロックすることができます。

- ・ AIガードレール機能を使うためには別途AIガードレールオプションを購入いただく必要があります。
- ・ 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。入力や出力内容は利用者で問題ないか確認が必要です。

2.2.2. カスタム認証（Active Directory）向け同期コマンドの提供

カスタム認証（Active Directory）を使用している場合に、Active Directoryのユーザ・グループを製品のユーザ・グループに同期するコマンドを利用できるようになります。ユーザのグループ所属情報も同期できます。コマンドはWindows端末上で動作でき、Windowsタスクスケジューラを用いることで同期の自動化も対応可能です。

本機能を使うためには追加のセットアップ作業が必要になります。詳細はカスタム認証セットアップガイドの「IdP連携(Active Directory)認証利用時の設定」を確認してください。

2.2.3. ユーザ・グループ情報のエクスポート機能の追加

登録されているユーザ・グループ情報をCSVファイルに出力する機能が追加されました。後述するインポート機能と合わせることでエクスポートした情報を編集して一括反映することができます。

2.2.4. ユーザ登録時の案内メール送付

ユーザ登録時に登録ユーザのメールアドレスに対して初回ログイン手順を示した案内メールを送付できるようになりました。本機能を使用するためには追加のセットアップ作業が必要になります。詳細はセットアップガイドの「認証方式の検討」を確認してください。

- ・ 本機能を使用するにはメールサーバと連携させる必要があります

2.2.5. 画像ファイル添付対話に対応

画像ファイルを添付して会話する機能が追加されました。チャット画面、推論APIで画像ファイルについてLLMと会話することができます。

本機能を使用するためには追加のセットアップ作業が必要になります。詳細はセットアップガイドの「拡張対話（ファイル添付）機能のセットアップ」を確認してください。

i 本機能を画像ファイルの入力に対応できるLLMモデルが使用できる外部LLMと連携させる必要があります

cotomi v3は画像ファイルの入力には対応していません

2.2.6. 回答のランダム性の上限値設定の追加

チャット画面の詳細設定にある回答のランダム性について、任意の上限値を設定できるようになりました。設定手順の詳細は運用ガイドの「回答のランダム性の上限を指定」を確認してください。また、上限値の初期値も1.0に変更になっています。

2.2.7. インデックスの登録文書のチャンク検索APIの追加

インデックスの登録文書に対して入力キーワードに関連するチャンク（文書の一部）を検索できるAPIを追加しました。

2.2.8. APIキー設定コマンドの追加

任意のAPIキーを設定したい場合などに使用できるAPIキーの設定コマンドを追加しました。

2.3. 機能強化

2.3.1. 管理ポータル機能の権限強化

管理者権限を管理機能ごとに細分化して設定できるようになりました。これにより例えばインデックス管理者とインデックスに文書登録する管理者を分け、文書管理のみ行う管理者を設定するなど権限譲渡ができるため、管理者の負担軽減が期待できます。

2.3.2. ユーザ・グループのインポート機能強化

従来のユーザの一括追加に加え、ユーザ情報の一括編集、グループの一括追加に対応しました。ユーザ情報はグループ所属情報も変更できるためグループ所属も一括で変更することができますようになりました。グループについてもCSVファイルからまとめて登録できる機能が追加されました。

2.3.3. 複数インデックス対話に対応

チャット画面にて複数インデックスを指定して対話できるように強化されました。従来では検索対話APIでのみ複数インデックスでの対話に対応しておりましたがチャット画面からも使用できるようになりました。

2.3.4. 複数ファイル、URLの添付対話に対応

複数ファイル、URLを添付して対話できるように強化されました。チャット画面、推論APIの両方で対応しています。

2.3.5. インデックス名の日本語対応

インデックスの名前に日本語が使用できるようになりました。

2.3.6. インデックスのユーザ権限対応

インデックスに対しグループではなくユーザ権限が付与できるようになりました。ユーザ単位でアクセス権が付与できるため、より柔軟なアクセス権制御に対応できるようになりました。

2.3.7. バックアップ・リストアによるロールバックに対応

従来バックアップ・リストアはサーバ故障などを想定した新規環境へのリストアのみに対応しておりましたが、ロールバック用途にも対応できるようになりました。これにより過去の時点のデータに戻すといったロールバック運用にも対応できるようになります。

2.3.8. 推論を用いた回答根拠の確認の対応モデル追加

推論を用いた回答根拠の確認の対応モデルにAzure OpenAI ServiceのGPT-4oだけでなく、OpenAI準拠のGPT-4oが追加されました。

2.3.9. 回答の最大トークン数の表示改善

チャット画面の詳細設定にある回答の最大トークン数について、値が未設定の場合に既定でどの値が使用されるのかを表示するように改善しました。

2.3.10. パスワードポリシーエラーのエラーメッセージ強化

ユーザ追加、変更時のエラーメッセージを強化し、パスワードポリシー準拠しないパスワード設定によるエラーであることが分かるようになりました。

2.4. 機能変更

2.4.1. 登録テンプレートの登録文字数上限の変更

テンプレート機能で登録できる文字数の上限が6万文字程度まで引き上げられました。

2.4.2. 拡張対話の文字数読み込み上限の撤廃

cotomi v3による入力文字数の上限引き上げを受け、拡張対話のファイル、URL添付時の文字数の読み込み上限が3000文字から上限が撤廃されました。

cotomi v3の入力文字数上限内であればファイルやURLにある文字数を気にせず会話することが可能になりました。

2.4.3. 検索対話のtopKオプションの上限引き上げ

検索対話のオプションの1つであるtopKの上限が100に変更になりました。

2.4.4. 回答の最大トークン数の初期値変更

チャット画面や対話API（Open AI API以外）の回答の最大トークン数の初期値が2048に変更になりました。すでに初期値から変更している場合は特に影響ありません（現在の設定値のままになります）

2.5. バグ対応

2.5.1. 会話履歴へのインデックス付与

会話履歴の保存データ数が増えていくとサーバのCPU、メモリの使用率が増えていく問題に対応しました。

2.5.2. カスタム認証連携時の管理ポータル画面でのログインユーザ操作制限

カスタム認証の外部IdP連携時にログインユーザの削除操作や役割の変更操作ができないように修正しました。

2.5.3. カスタム認証(EntraID)連携時のユーザの表示名に日本語などの特定文字があるとログインできない

カスタム認証の外部IdP（Entra ID）連携時にEntra ID側のユーザの表示名に日本語があると画面にログインできない問題に対処しました。ログイン後、ユーザの名前は文字化けすることがありますが、操作は問題ありません。文字化けについても今後改善予定です。

参考：バージョンの確認方法

対応バージョンの確認方法は以下の通りです。

i HTTPSに既定では自己証明書を使用しています。その場合「この接続ではプライバシーが保護されません」と注意が表示されますが、問題ありませんので接続してください。

- チャット画面

1. Webブラウザを立ち上げ以下にアクセスします。アクセス先はGENAI_DOMAINで指定した値になります。

 https://<GENAI_DOMAINの指定値>/

2. 画面右下に記載しているbuild番号を確認します（以下の赤枠）。



- 管理ポータル

1. Webブラウザを立ち上げ以下にアクセスします。アクセス先はGENAI_DOMAINで指定した値になります。

 https://<GENAI_DOMAINの指定値>/admin

2. 画面右下に記載しているbuild番号を確認します（以下の赤枠）。

Generative AI 管理ポータル

[ログイン](#)

