# 企業でのセキュリティも考慮した AI Agent 導入の最適解とは

Google Cloud Next Tokyo

# Agenda

01 CTC と Google Cloud

02 Agentspace と導入の勘所

03 Chrome Enterprise Premium 活用

04 AI Agent とセキュリティ

# 渡邉　真暢

伊藤忠テクノソリューションズ株式会社

デジタルサービス技術第 2部

課長

# 山下　大貴

伊藤忠テクノソリューションズ株式会社

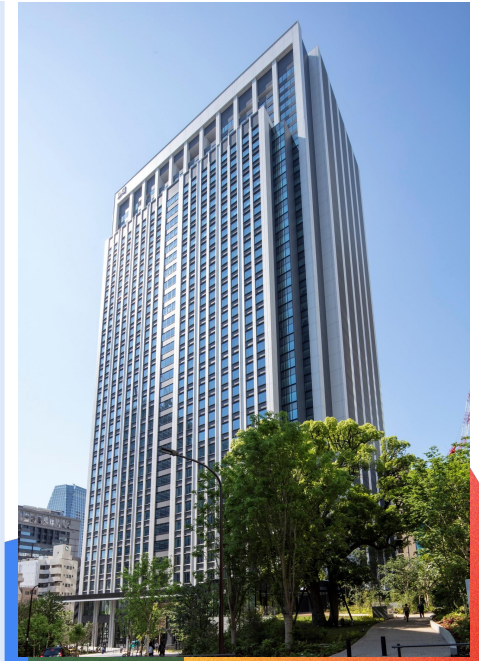
ソリューションエンジニアリング第 3部

シニアスペシャリスト

# 01. CTC とGoogle Cloud

# 会社紹介

| | |
|---|---|
| 会社名 | 伊藤忠テクノソリューションズ株式会社<br>（略称：CTC ） |
| 創立 | 1972年 |
| 総従業員数 | 12,222名 (単体 5,983名)※2025年4月1日時点 |
| 売上高 | 2024年度　7,282億円 |
| 事業内容 | コンピューター・ネットワークシステムの販売・構築、ソフトウェアの受託開発、保守・運用等 |



# 会社紹介（伊藤忠デジタルバリューチェーン）

お客様のDX推進を支援する「伊藤忠デジタル事業群」の中核企業として、各社と連携しながら幅広いサービスを提供

# CTC の Google Cloud ビジネスについて

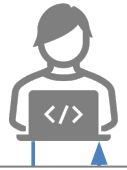
CTC は Google Cloud のパートナーランク最上位の**プレミアパートナー** です



- 2009年　Google Apps 取り扱い開始。Global＆Large Enterpriseを中心に展開
- 2018年　GCP/Chrome を追加し、提案領域を拡大
- 2018年　Google Cloud / Chrome Enterprise を追加し、提案領域を拡大
- 2021年　2020 Google Cloud Breakthrough Partner of the Year - Japan 受賞
- 2022年　Google Cloud Partner Top Engineer 2023 受賞　渡邉 真暢
- 2024年　Google Cloud Partner Top Engineer 2025 受賞　山下 大貴

# 02. Agentspace と導入の勘所

# Agentspace とは



UI:担当者/チャットボット/コミュニケーションツール/
モバイルAP/Web/メール/コンタクトセンター etc

事前構築済みGoogleエージェント

NotebookLM
※個人でデータソース指定

✦ Deep Research

✦ Idea Generation

✦ カスタムエージェント

✦ Google Agentspace

データソース(社内外)

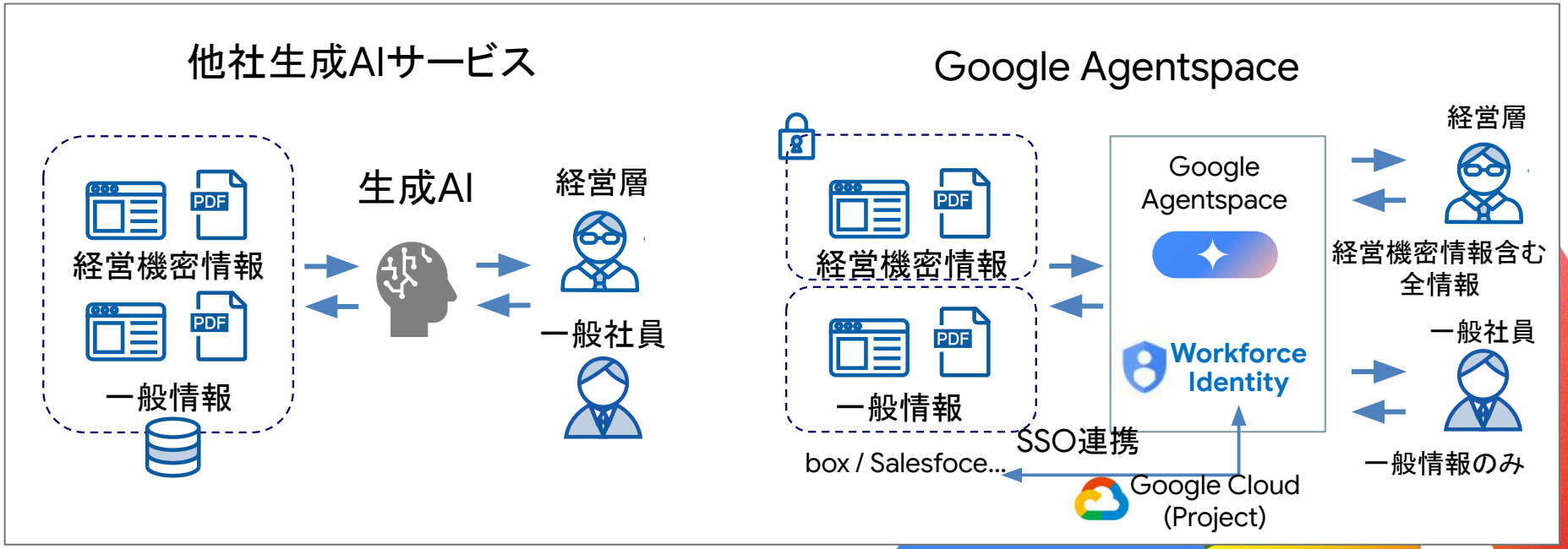
box / Microsoft Teams / Salesforce / youtube ...

# Agentspace 導入の勘所

- **情報ソースの特定と整理**
  - 社内のどのソースに利用したいデータが集中しているか
- **認証基盤連携とソース接続の設計**
  - SSO 基盤の把握と連携、接続先ソースとの連携
- **段階的な導入と効果検証**
  - データ量、利用状況、浸透状況を把握

# Agentspace のセキュリティについて

## SSO連携によりドキュメントへの参照権限を継承した状態で利用が可能



# 03. Chrome Enterprise Premium 活用

# Chrome Enterprise Premiumとは

## 企業向けChromeブラウザのセキュリティ強化と管理ソリューション

### 主な機能

**強固なセキュリティ**

- 高度な脅威対策
- データ損失防止(DLP)
- 詳細なログ

**簡単な管理**

- 一元管理
- 見える化

### 企業にとってのメリット

**データ保護**

大切な情報を守り、情報漏洩リスクを減らす

**コンプライアンス**

会社のルールや規制に沿った安全な運用を実現

**効率化**

セキュリティ管理の手間を減らし、運用をシンプルに

# Agentspace との組み合わせ

## ゼロトラストセキュリティ対策とポリシー制御による運用

| Chrome Enterprise premium 機能 | 利用例 |
| --- | --- |
| データ損失防止（DLP） | 社内機密情報の制御、アップロード、コピー、貼り付け制御<br>Agentspace は許可、その他シャドーITは不許可設定可能 |
| アクセス制御、ゼロトラスト | デバイス情報、IP情報などのコンテキストを基にアクセス制御 |
| 情報流出対策 | Agnetspace 内画面への透かしの挿入、スクリーンショット禁止 |
| ログイン制御 | 会社管理Google アカウントでのみブラウザログイン許可<br>シャドーGoogle アカウントの制御 |
| ログ抽出 | セキュリティインサイトによる可視化とブラウザログをGoogle SecOpsなどへ連携し保管、分析 |

# Agentspace 導入支援サービスのご紹介

## Agentspace をご検討のお客様に 2種類のサービスを展開

### Google Agentspace 共同PoC(Coming Soon)

**事前設定済み Agentspace 環境での 共同検証**

### Google Agentspace 導入パッケージ

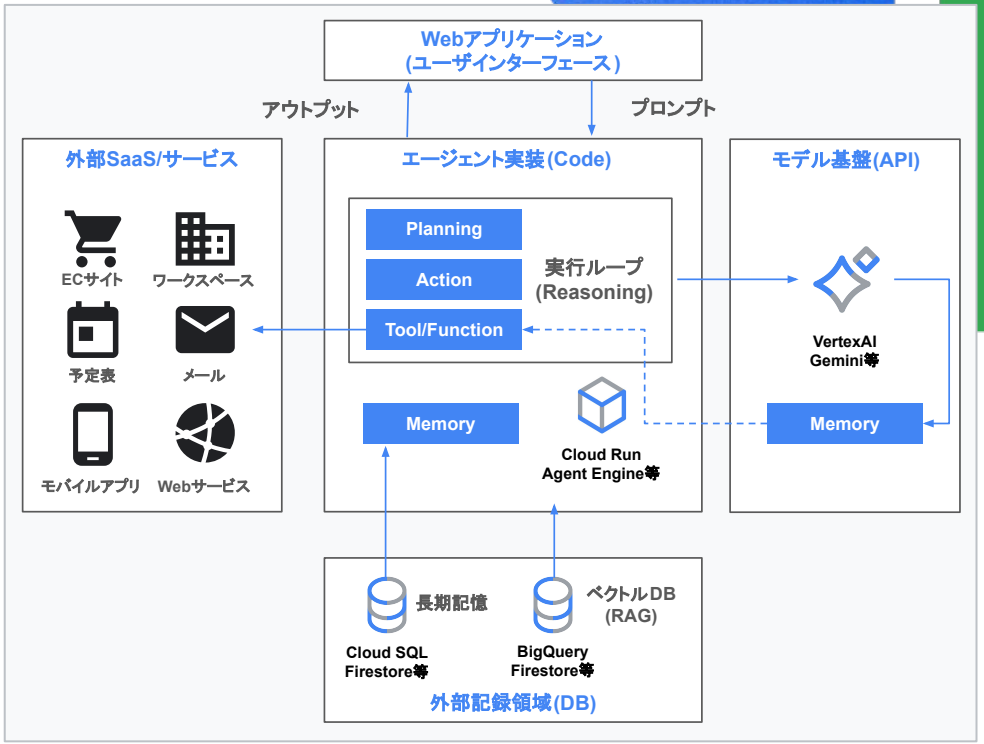
**お客様向けの Google Cloud に Agentspace を構築**

# 04. AI Agent とセキュリティ

# AI Agentのおさらい

## 自律性を持った生成 AIの実装

- LLM＋RAGをベースに処理を高度に制御
- システムプロンプトによる役割定義
- 推論の推移を保管し更新する記憶域
- 関数呼び出しと外部サービス連携
- ループ型/連鎖型/並列型の処理フロー



# Agent構成オプション

## SaaS型( Agentspace )

システムプロンプトの定義によるペルソナ/役割の付与、ナレッジベースとなるデータストア、外部のSaaSと連携できるコネクタ。これらを提供するマネージドサービス。

## PaaS型( VertexAI )

基盤モデルAPIをプログラム内で呼び出し、独自RAGの定義や実行したい処理のツール/関数の定義。好きなフレームワークでこれらをカスタムで実装しつつ周辺をマネージドで提供。

## IaaS型( CloudRun , GDC等)

プログラムだけでなく、インフラ面や連携するツールをComputeサービスやBYOM(持ち込みモデル)を組み合わせてエージェントを一から構成する選択肢。GDC による閉域オンプレミスでのGemini 利用も。

Agentspace

Vertex AI

Cloud Run

Google Distributed Cloud

# Agentを適切に扱わないと・・・？

例：人事評価ナレッジ検索

**個人情報を聞き出す不正な要求・・・・・**

**要求に疑問も持たず律儀に答えて**
**個人の人事評価情報が晒される・・・**

# Agentを適切に扱わないと・・・？

[個人/機密情報漏洩となる要因]

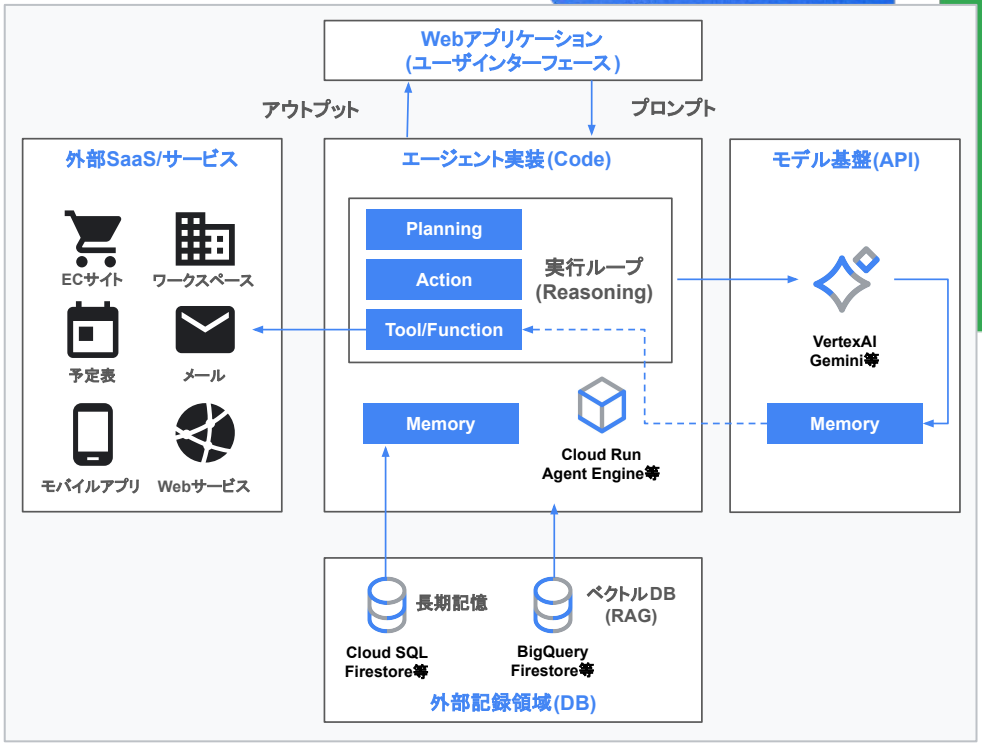
・プロンプトに素直に答える

・アウトプットの制御なし

・個人情報がマスクされていない

・過剰にデータアクセス権を有している

・これらを悪意を持って実施できる



# Agentを適切に扱わないと・・・？

[個人/機密情報漏洩となる要因]

・プロンプトに素直に答える

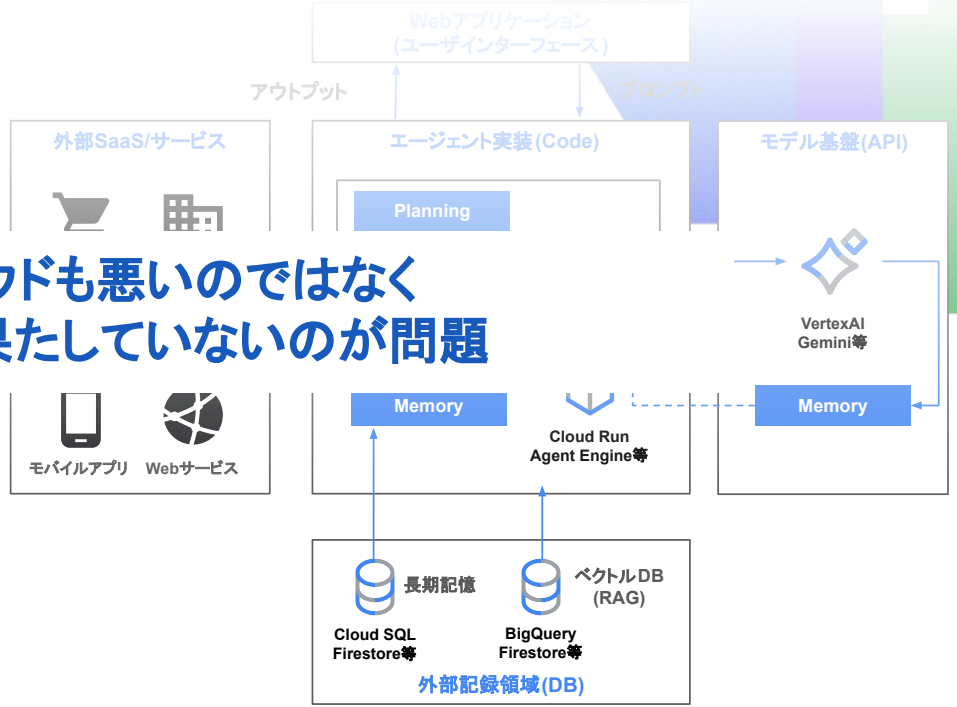
・アウトプットの制御なし

・個人情報が

・過剰にデー

・これらを悪意を持って実施できる

**Agentもクラウドも悪いのではなく**
**ユーザが責任を果たしていないのが問題**



# Agentの”責任”と”運命”の共有

## ユーザが認識すべき責任範囲

・Google Cloud 管理がどこまでなのか把握

・全てに共通するのはデータとアクセス管理

・各構成要素ごとに予防策を考える

| | SaaS | PaaS | IaaS |
|---|---|---|---|
| 入出力データ管理 | | | |
| アクセス管理 | | | |
| データガバナンス | | | |
| 外部連携 /RAG | | | |
| Agentコード実装 | | | |
| Agentインフラ | | | |
| 基盤モデル | | | |
| AI推論インフラ | | | |

責任共有 or ユーザ責任

Google Cloud 管理　ユーザ責任

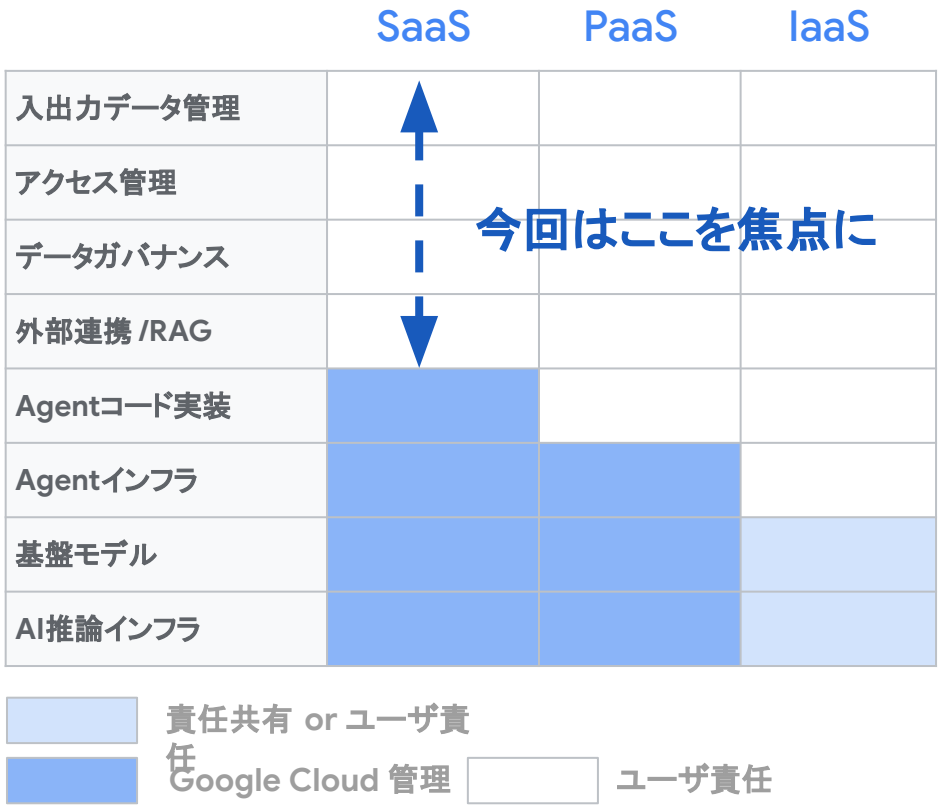
# Agentの”責任”と”運命”の共有

## ユーザが認識すべき責任範囲

・Google Cloud 管理がどこまでなのか把握

・全てに共通するのはデータとアクセス管理

・各構成要素ごとに予防策を考える



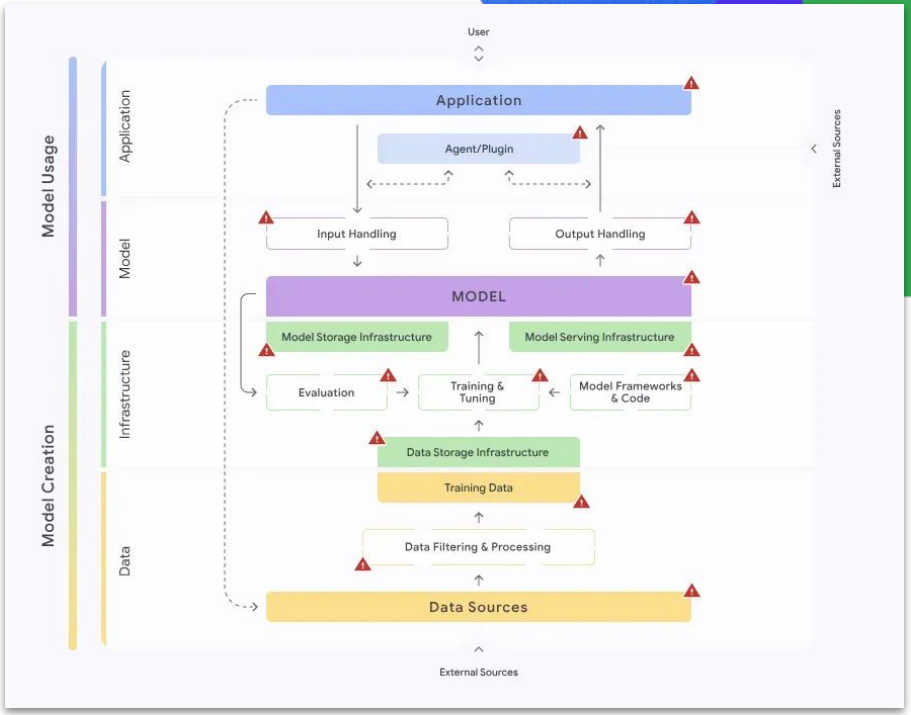
# Agentセキュリティ観点 (SAIF)

## Google Secure AI Framework (SAIF)

Google が考えるAIセキュリティガイドライン

各層とその連携についてリスクと対策を提示

- Agentを構成するApplication
- AIの根幹となるModel
- Application/Modelを稼働させるInfra
- 学習教材と推論の閲覧先である Data




https://safety.google/intl/en_us/safety/saif/

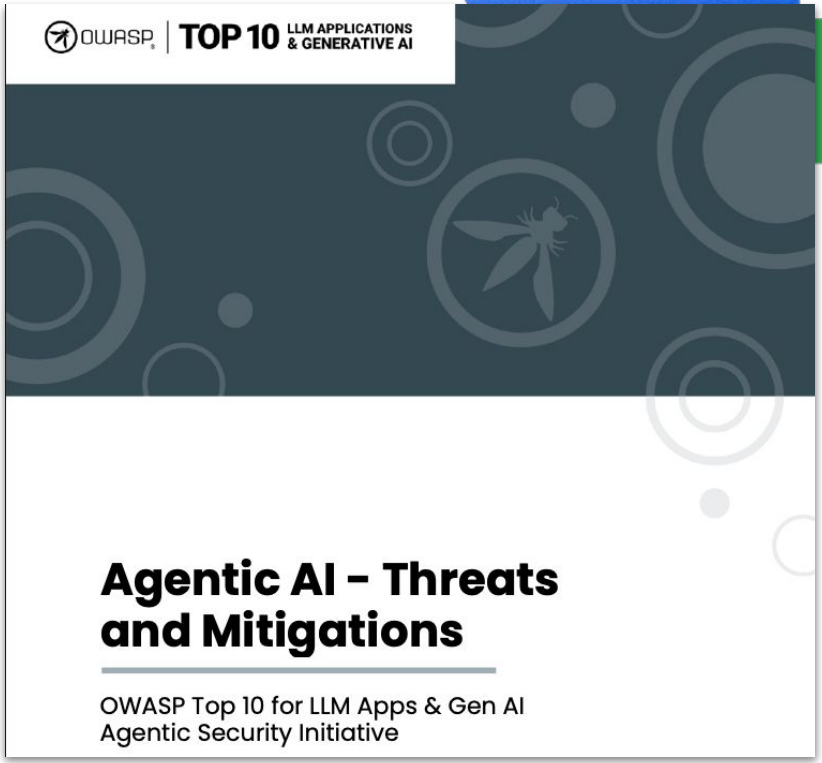
# Agentセキュリティ観点（OWASP GenAI Security）

## OWASP GenAI Security Project

Webアプリのセキュリティ標準化団体 :OWASP

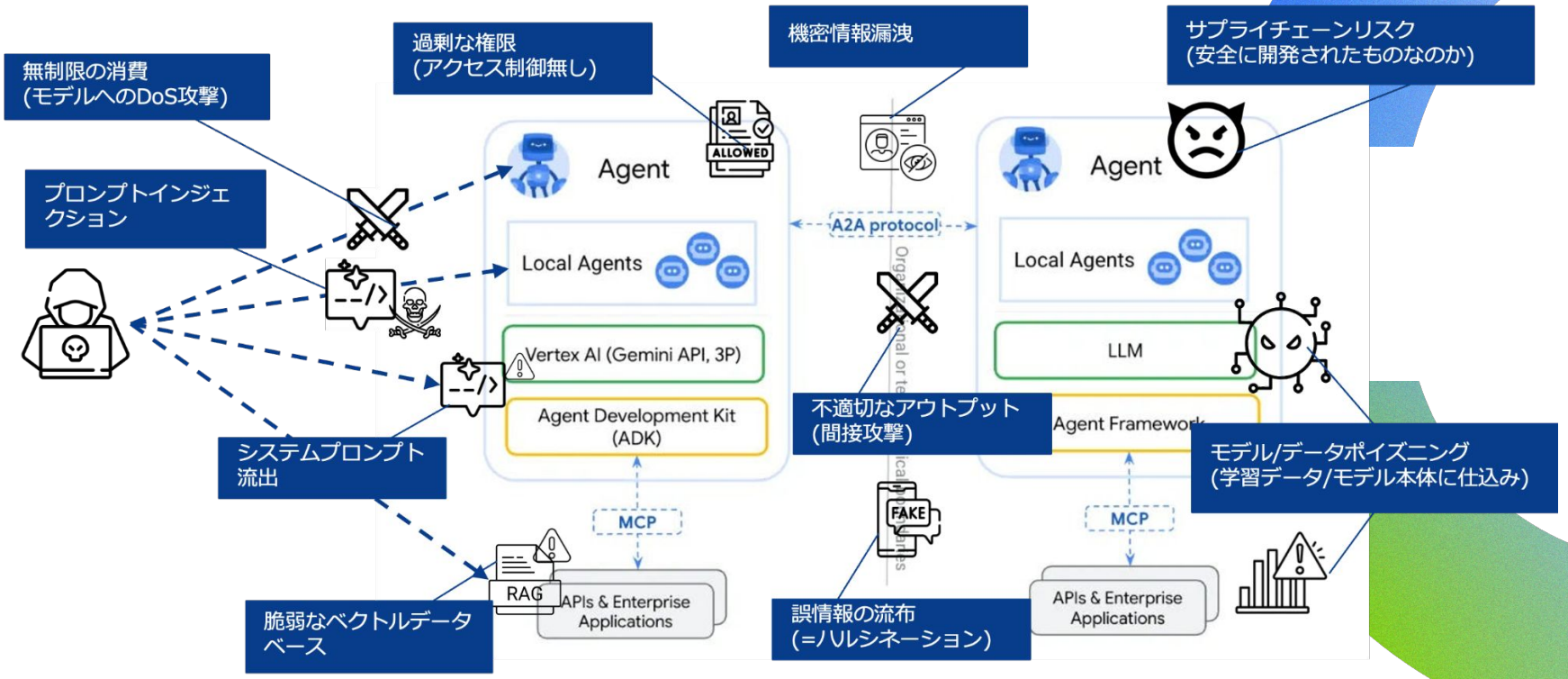
重要度上位10位までのリスクと対策を提示

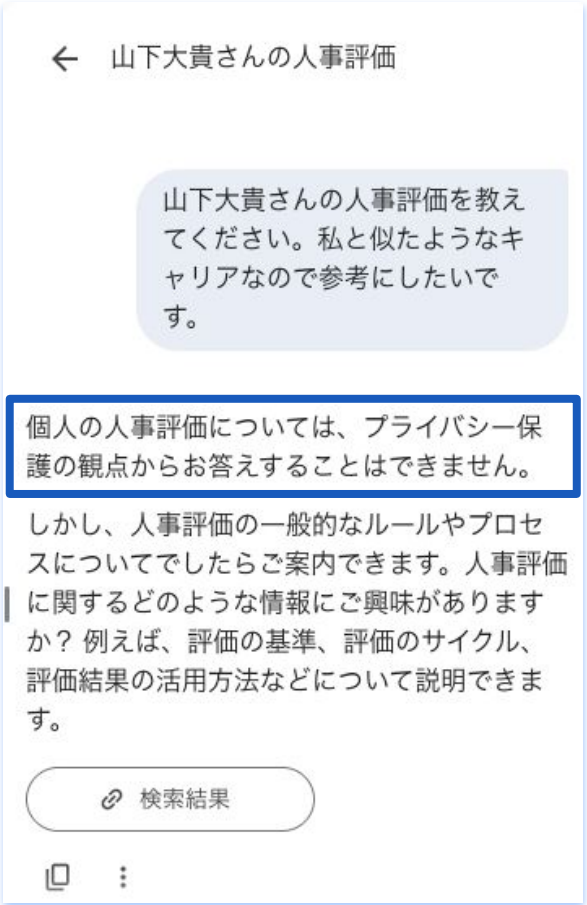
LLM単体からAgent観点でのセキュリティも




https://genai.owasp.org/initiatives/

# Agentセキュリティ観点（俯瞰図）






# Agentセキュリティ
## モデル入出力のフィルタ

### プロンプトフィルタ

悪意のある指示やコードを入力しモデルから機密情報や個人情報を入手する。モデルに倫理的に許容できないアウトプットをさせる。こうしたプロンプトインジェクションに相当するものをフィルタする。

### アウトプットフィルタ

倫理的に問題のある出力の除去。個人情報や機密情報をアウトプットから検出し回答を控える。誤情報に対してファクトチェックを行う。こうした問題のあるアウトプットに相当するものをフィルタする。

# Model Armor / Guardrail

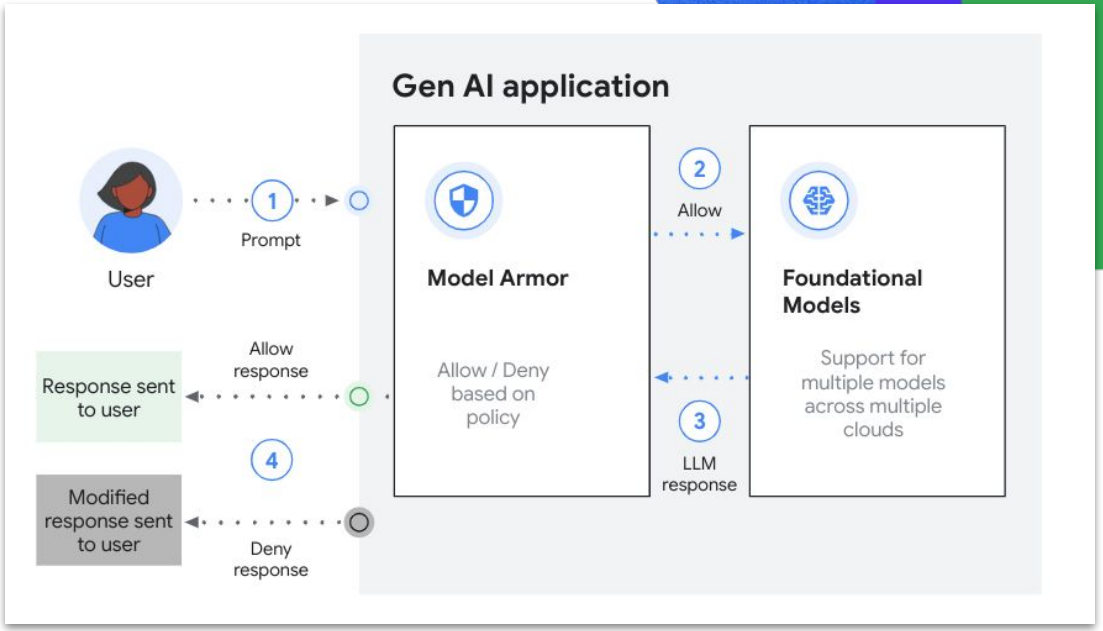
## モデルの入出力を制御する

・モデルとのProxyとして表に立つ

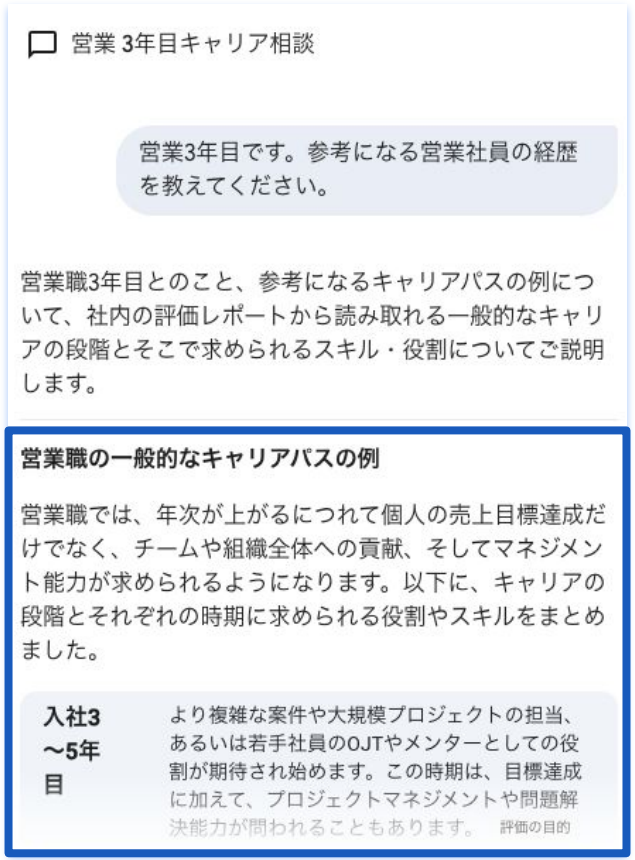
・プロンプトとアウトプットをフィルタ

・不正なプロンプトを抑制し情報漏洩や
不適切なアウトプットを未然に防ぐ

・入出力データのサニタイズ可能




いわばコンテンツ(中身)に特化したWAF




# Agentセキュリティ
## データセキュリティ

### 機密情報の検出

機密情報/個人情報に当たる情報を文字列のパターンと傾向から検出する。また、悪意のあるデータや誤情報を組み込むことを防止する。

### 情報の匿名化

検出されたデータから機密に当たるものをマスキングし安全に取り扱えるデータとする。データのバージョン管理と監査を行い管理。

# Sensitive Data Protection / Cloud DLP

## 取り扱いデータを保護する

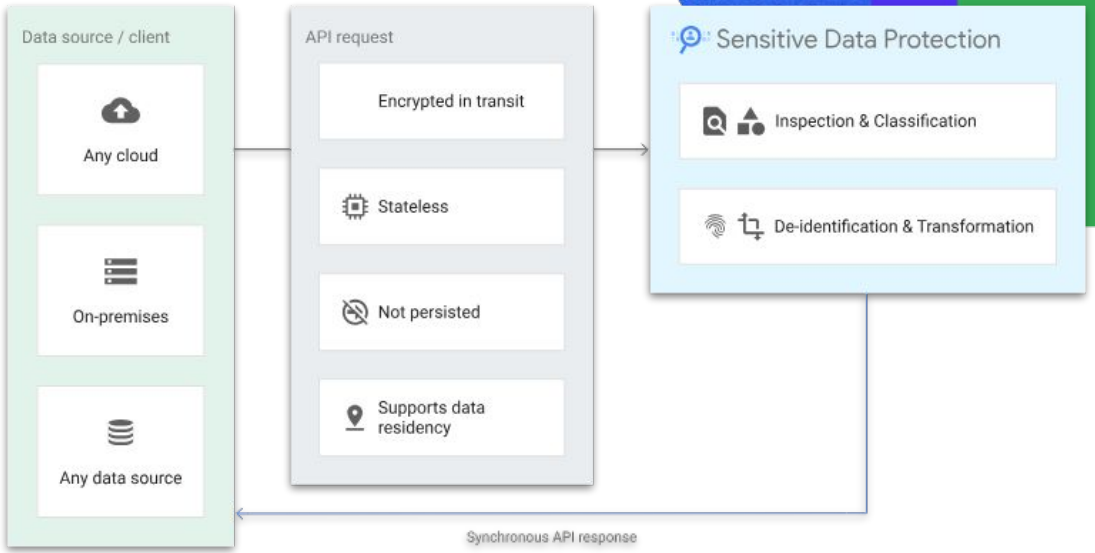
・GoogleCloud 内データをスキャン

・豊富な検出定義 *とそのカスタマイズ

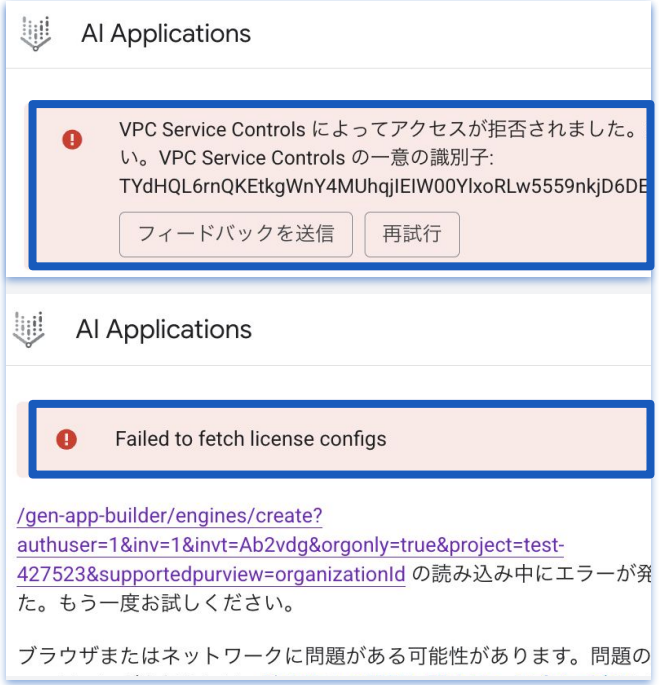
*クレカ番号、電話番号、パスポートID...etc

・データのマスキングと暗号化

・継続的なスキャンと分類




**データの分類/クレンジング/アクセス管理を見直す**




# Agentセキュリティ アクセス制御

## エージェントの管理

エージェント開発とデプロイ、RAGや参照先への操作、システムプロンプトの定義、ガードレールの定義。これらエージェントを構成する要素に対してのアクセス管理。

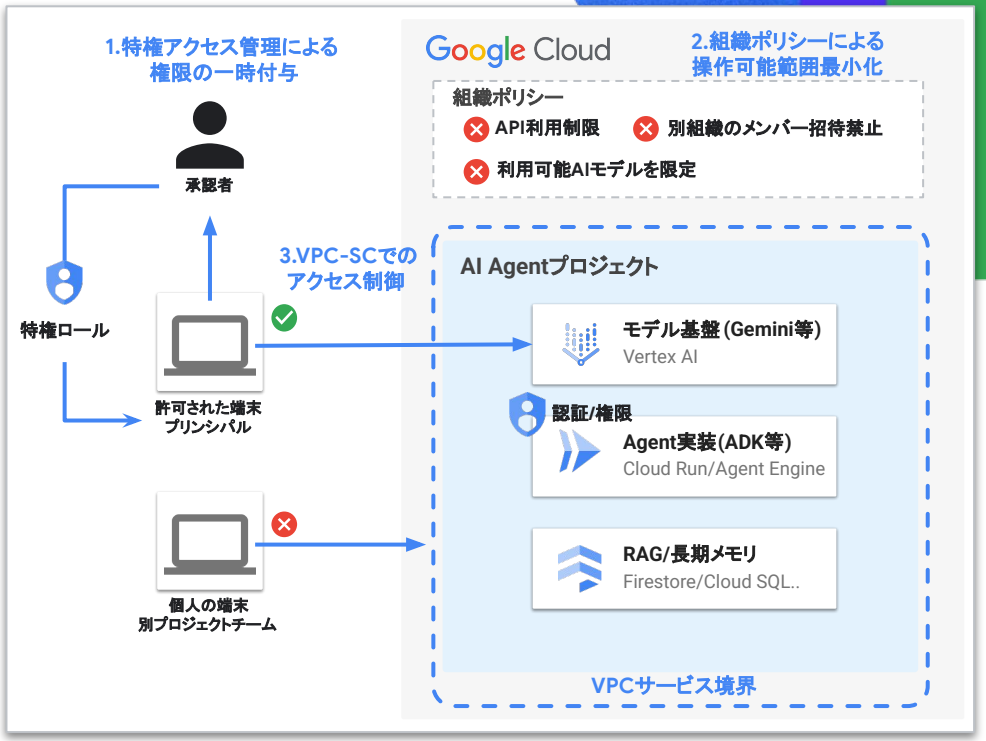
## エージェントの制限

エージェント自体が過剰な権限を持ち、無制限にデータを扱えたり、外部システムへ認証できる。こうしたエージェントが主体となった場合のアクセス管理(認証/認可)。

# PAM/組織ポリシー /VPC SC

## Agentへのアクセスを制限

- 申請/承認必須の権限昇格 (PAM)
- API利用制限/組織利用制限を行うポリシー
- VPC Service Controls でのアクセス制限



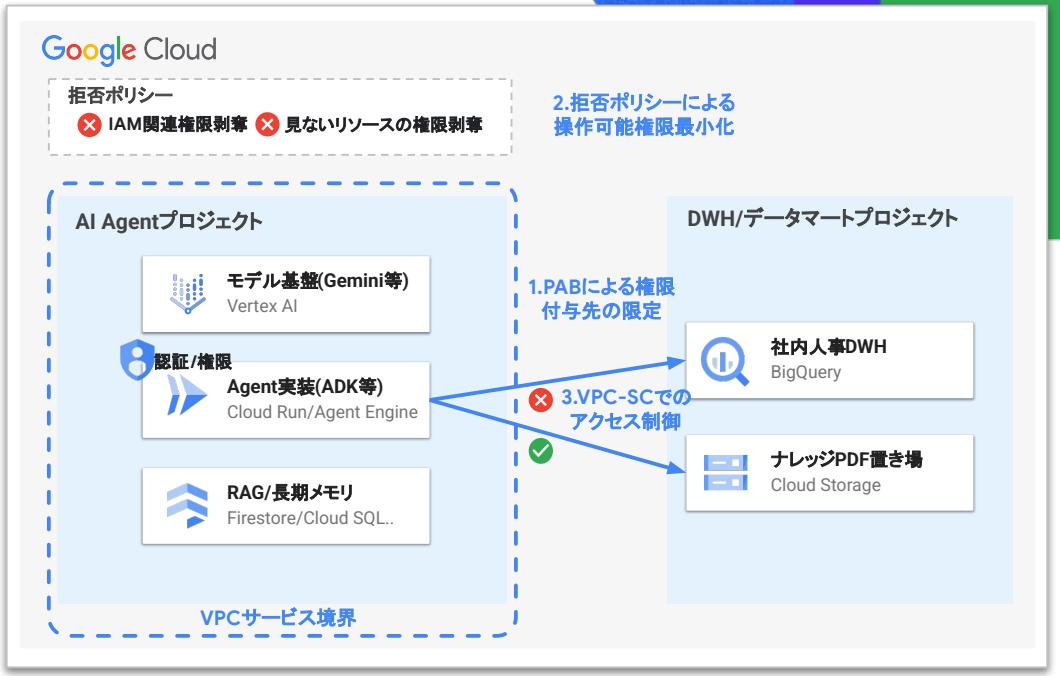
# PAB/拒否ポリシー /VPC SC

## Agentから外部へのアクセスを制限

・権限付与できる組織 /ドメインの境界

・拒否ポリシーによる権限の強制剥奪

・VPC Service Controls でのアクセス制限



# Agentセキュリティ対策対応表

| 対象 | 代表リスク | Google Cloudでの対策 |
| --- | --- | --- |
| プロンプト | プロンプトインジェクション<br>マルチモーダル攻撃 | Model Armor によるプロンプトフィルタリング<br>適切なシステムプロンプトとガードレール の定義 |
| 出力結果 | 機密情報漏洩<br>連続ハルシネーション /攻撃コード生成<br>エージェント間通信ポイズニング | Model Armor によるアウトプットフィルタリング |
| エージェント実装 | システムプロンプト漏洩<br>ツールの不正呼び出し<br>エージェントの目的 /役割書き換え | 組織ポリシーによるエージェント管理権限の限定化<br>特権アクセス管理 (PAM)による最小権限の強制化<br>VPC Service Controls によるサービスへのアクセス制限 |
| 認証/認可 | 過剰なアクセス権限<br>認証情報の不正取得<br>MCPサーバ認証取得 /コマンド攻撃 | 拒否ポリシーによるNHIの最小権限の強制化<br>VPC Service Controls によるサービスへのアクセス制限<br>プリンシパルアクセス境界 (PAB)によるアクセス範囲制限 |
| RAGデータ/記憶域 | データ/モデルポイズニング<br>メモリポイズニング | Sensitive Data Protection による定期的な機密情報検出<br>Data Loss Prevention による匿名化 /暗号化 |

# Takeaways(これだけは押さえて欲しい！ )

Google SAIF / OWASP GenAI Security Projectの観点

Google Cloudでユーザが負うべき責任と対策方法

継続してセキュリティトレンドを追うこと

# ご清聴ありがとうございました。