Engineer

運用中のAIエージェントのセキュリティ

AIエージェントのアクセスをレビューし、制御を実装します。合意した対象一覧、優先順位付きの指摘、権限テスト、検証記録を提供します。

AIエージェントが何にアクセスできるかを把握し、そのアクセスを制御します。メール、文書、APIに接続するエージェントには、確認すべき権限と、問題が生じる経路が加わります。合意したシステムを一覧にし、リスクを評価し、承認された制御を実装して、制限した操作をテストします。アクセスの構成図、指摘事項、検証記録をお渡しします。

サービスを成果につなげる

「運用中のAIエージェントのセキュリティ」を事業の成果につなげましょう。技術的な作業を、それが支えるべき利用者、運用、意思決定に結び付けるお手伝いをします。 各エージェントのアクセスを必要な範囲に限定し、エージェント自身が改変できない場所に操作を記録し、安全策を信じるだけでなく検証します。

サービスを見る お問い合わせ →

関連する実績を確かめる

このページの関連実績から、私たちの取り組み方をご確認ください。それぞれの成果の背景にある条件や、何を優先して何を見送ったかについてもお尋ねください。 Linuxサーバーを守る1,000件のRBACルールを作成 — 1,000 セキュリティ上の穴を塞ぐアクセスルール

Linuxサーバーを守る1,000件のRBACルールを作成 お問い合わせ →

把握する：各エージェントとアクセスできるもの

合意した範囲内のエージェントを一覧にします。承認済みの連携と、従業員の方々と確認したその他の接続も含めます。構成図には、それぞれの認証情報、ツール、到達可能なデータを記載します。指摘事項とともに構成図をお渡しし、最新に保つ責任を引き継ぎの一部として合意します。これは、レビューした環境の記録であり、対象範囲と未確認の部分を明示します。

制限する：各エージェントに最小限の権限を

エージェントには固有のIDと、業務に必要な最小限の権限を与えます。共有アカウントではなくロール（RBAC）を使い、原則として読み取り専用にし、書き込みは頻度と量を制限した、用途の狭いツールだけを通します。シークレットはエージェントごと、ホストごとに分け、プロンプトには決して貼り付けず、コードを変えずに更新します。メール、ウェブページ、文書などの外部の文章は、信頼できる命令ではなく、検証が必要なデータとして扱います。ツールの権限と検査によって合意した境界を設け、「ルールを無視せよ」というメッセージを操作の許可とは見なしません。検査が実行できなければ操作を拒否し、安全側に倒します。

記録する：何を、誰のために行ったか

すべてのツール呼び出しについて、誰の代理で動き、何を求められ、何に触れ、何が返ったかを記録し、エージェント自身が変更できない場所にログを保持します。システムに応じてアラートのルールを合意します。読み取りの急増、想定時間外の利用、エージェント用に合意したデータ範囲を超えるアクセスなどを対象にできます。実際に問題が起きたとき、その記録によって、インシデントを推測ではなく時系列で把握できます。

構成図も、制限も、ログも **三つの主張にすぎません** — 誰かがそれを突破しようと試すまでは。

合意した制御は、代表的なアクセス、不正利用、復旧の事例で、引き継ぎ前に検証します。

制限は信じるだけでなく試す

契約の範囲内でアクセス制御をテストし、何を試し、何が拒否され、何に引き続き対処が必要かを記録します。テストには、拒否すべきツール呼び出し、レート制限、認証情報の分離、プロンプトインジェクションの試行を含められます。成果物は、確認可能な検証記録と、限界を明示した再実行可能な検査一式です。すでにある漏えいの封じ込めが必要な場合、インシデント対応と認証情報の更新は別途範囲を定めます。

取り組みを裏付ける実績

サーバー、ID管理システム、プラットフォームで、この仕事を行ってきました。Linuxとデータベースサーバーにまたがる1,000のRBACルールを記述し、M365とシングルサインオン全体のアクセス管理を主導し、21の文書化された監査を伴う継続的なセキュリティスキャンを運営しました。運用上のすべてのシークレットをホストごとに分離し、一秒で行うシークレットの事前検査を追加し、三つの防御策がどれも機能していないことを発見し、侵入元の遮断が毎回機能することを検証しました。検査不能時に許可しないレート制限を構築し、漏えい後にシークレットの適用範囲のルールを作成しました。自社での取り組みは、セキュリティへの取り組みに記載しています。

ご依頼の進め方

まず、変更を加えずに調査します。合意した範囲内のエージェント、その認証情報、アクセス先を整理し、理由と修正方法を添えた指摘事項をリスク順にお渡しします。その後、お客様の設定とリポジトリで、お客様のレビューを通して修正します。対象のアクセス経路で、合意した権限を超えたら失敗する検査も備えます。構成図、記録、テストはお客様の手元に残り、監査人や保険会社に渡す資料になります。エージェントが読むデータにも、同じように権限を設けられます。他の仕事と同じく、料金に記載する条件でご依頼いただけます。

すでにエージェントが社内システムへの鍵を持っている場合は、
対象をお聞かせください。

対象のAIエージェントのセキュリティレビュー

把握する

対象とするエージェント、認証情報、データへのアクセスを一覧にして合意し、設定とともに管理します。

制限する

各エージェントに固有のIDと最小限の権限を与え、読み込む文章は命令ではなくデータとして扱います。

記録する

誰のために行ったかを含め、すべての操作を、エージェントが変更できない場所に記録し、通常と異なる動作には通知を出します。

自社のエージェントについて相談する

運用するエージェントはすべて **何かにアクセスする鍵を持っています** — どの鍵で何に届くのかを把握しましょう。

まず、変更を加えずに、対象のエージェントと認証情報の対応関係を整理し、リスク順に指摘事項をまとめます。

エージェントの構成調査を相談する 料金 →

発注前に業務範囲を定める

デプロイの流れ、運用上のリスク、緊急対応の担当者をお知らせください。検査、復旧手段、明確な責任の所在を伴う変更の範囲を定められます。

料金を見る お問い合わせ →

要件について話し合う

「運用中のAIエージェントのセキュリティ」に求めることを、納期や保守を担当するチームも含めてお聞かせください。実行可能な進め方をご相談いただけます。

サービスを見る

welcome@engineer.company +45 50 10 61 84 デンマーク、コペンハーゲン

このページを持ち歩く

連絡先カード

心を込めて丁寧に制作： D.

CVR 45969703 · Tagensvej 105C, 3. th, 2200 København N

著作権 © 2025 – 現在 · Engineer ApS · すべての権利を留保します。

印刷元： https://engineer.company/ja/ai-security/