Engineer

Sicherheit für Ihre KI-Agenten

Zugriffsprüfung und Kontrollen für KI-Agenten: vereinbarter Bestand, priorisierte Befunde, Berechtigungstests und dokumentierte Verifikation.

Verstehen Sie, was Ihre KI-Agenten erreichen können, und sichern Sie diesen Zugriff ab. Die Anbindung an E-Mails, Dokumente oder eine API schafft Berechtigungen und mögliche Fehlerpfade, die geprüft werden müssen. Wir erfassen die vereinbarten Systeme, bewerten Risiken, setzen freigegebene Kontrollen um und testen die beschränkten Aktionen. Sie erhalten die Zugriffsübersicht, die Befunde und die Verifikationsnachweise.

Die Leistung in ein Ergebnis überführen

Machen Sie „Sicherheit für Ihre KI-Agenten“ für ein Geschäftsergebnis nutzbar. Wir können die technische Arbeit mit den Nutzern, Betriebsabläufen und Entscheidungen verbinden, die sie unterstützen soll. Der Zugriff jedes Agenten wird erfasst und auf das Nötige begrenzt, jede Handlung außerhalb seiner Änderungsbefugnis protokolliert und jede Schutzmaßnahme geprüft statt bloß vorausgesetzt.

Unsere Leistungen entdecken Kontakt aufnehmen →

Die passenden Arbeiten prüfen

Prüfen Sie unser Vorgehen anhand der zugehörigen Arbeiten auf dieser Seite. Fragen Sie nach den Abwägungen und den Bedingungen hinter jedem Ergebnis.

Fallstudien Kontakt aufnehmen →

Erfasst: jeder Agent und sein Zugriffsbereich

Wir erfassen die Agenten im vereinbarten Umfang, einschließlich genehmigter Integrationen und weiterer mit Ihren Mitarbeitern ermittelter Verbindungen. Für jeden hält die Übersicht Zugangsdaten, Werkzeuge und erreichbare Daten fest. Sie erhalten sie zusammen mit den Befunden; die Zuständigkeit für ihre Aktualisierung wird bei der Übergabe vereinbart. Sie dokumentiert die geprüfte Umgebung und weist ihren Umfang und ihre Lücken ausdrücklich aus.

Begrenzt: minimale Rechte für jeden Agenten

Ein Agent erhält eine eigene Identität und die kleinsten für seine Aufgabe nötigen Rechte: Rollen statt gemeinsamer Konten (RBAC), standardmäßig nur Lesezugriff und Schreibzugriffe ausschließlich über eng begrenzte Werkzeuge mit Häufigkeits- und Mengenlimits. Geheimnisse werden je Agent und Host getrennt, nie in einen Prompt eingefügt und ohne Codeänderung rotiert. Externe Texte, die ein Agent liest — E-Mails, Webseiten oder Dokumente —, werden als nicht vertrauenswürdige Daten behandelt. Werkzeugberechtigungen und Validierungsprüfungen setzen die vereinbarten Grenzen durch; eine Nachricht wie „Ignoriere deine Regeln“ ist keine Erlaubnis für eine Aktion. Die Kontrollen verweigern im Zweifel den Zugriff: Kann eine Prüfung nicht ausgeführt werden, wird die Aktion abgelehnt.

Dokumentiert: was der Agent tat und für wen

Jeder Werkzeugaufruf wird mit der Person protokolliert, für die er erfolgte, sowie mit Anfrage, betroffenen Daten und Ergebnis. Das Protokoll liegt an einem Ort, den der Agent selbst nicht verändern kann. Die Alarmregeln werden für das System vereinbart. Sie können einen plötzlichen Anstieg von Lesezugriffen, eine Nutzung außerhalb der vorgesehenen Zeiten oder Zugriffe über den vereinbarten Datenumfang des Agenten hinaus abdecken. Geht etwas schief, macht die Dokumentation aus einem Vorfall einen zeitlich nachvollziehbaren Ablauf statt einer Vermutung.

Eine Übersicht, ein Schloss und ein Protokoll **sind drei Behauptungen** — bis jemand versucht, sie zu umgehen.

Die vereinbarten Kontrollen werden vor der Übergabe anhand repräsentativer Zugriffs-, Missbrauchs- und Wiederherstellungsfälle geprüft.

Zugriffssperren testen statt voraussetzen

Wir testen die Zugriffskontrollen innerhalb des Auftragsumfangs und dokumentieren, was versucht und verweigert wurde und wo noch Handlungsbedarf besteht. Die Tests können abgelehnte Werkzeugaufrufe, Ratenbegrenzungen, die Trennung von Zugangsdaten und Prompt-Injection-

Versuche abdecken. Das Ergebnis sind prüfbare Verifikationsnachweise und wiederholbare Kontrollen mit ausdrücklich benannten Grenzen. Müssen bestehende Offenlegungen eingedämmt werden, werden Vorfallsbearbeitung und Rotation von Zugangsdaten gesondert abgegrenzt.

Die Arbeit dahinter

Wir haben diese Arbeit an Servern, Identitätssystemen und Plattformen geleistet. Wir haben 1.000 RBAC-Regeln für Linux- und Datenbankserver geschrieben, das Zugriffsmanagement über M365 und Single Sign-on hinweg geleitet und ein kontinuierliches Sicherheitsprüfprogramm mit 21 schriftlichen Audits betrieben. Wir haben alle sensiblen Zugangsdaten für den Betrieb je Host getrennt, eine Vorabprüfung für Geheimnisse mit einer Sekunde Laufzeit ergänzt, drei Schutzmechanismen gefunden, von denen keiner griff, und die Wirksamkeit der Angriffssperre bei jedem Lauf nachgewiesen. Wir haben eine im Fehlerfall sperrende Ratenbegrenzung aufgebaut und nach einem Zugangsdatenleck eine Regel zum Geltungsbereich geschrieben. Wie wir selbst arbeiten, steht unter So arbeiten wir sicher.

So läuft ein Auftrag ab

Am Anfang steht eine Prüfung mit reinem Lesezugriff. Wir erfassen die Agenten im vereinbarten Umfang, ihre Zugangsdaten und deren Reichweite und übergeben Ihnen die Befunde nach Risiko geordnet, jeweils mit Begründung und Abhilfe. Die Korrekturen erfolgen anschließend in Ihrer Konfiguration und Ihrem Repository und durch Ihre Prüfung. Kontrollen schlagen fehl, wenn die abgedeckten Zugriffswege die vereinbarten Berechtigungen überschreiten. Sie behalten Übersicht, Nachweise und Tests — die Unterlagen für Prüfer oder Versicherer. Die von Agenten gelesenen Daten können auf dieselbe Weise mit Berechtigungen versehen werden. Die Beauftragung erfolgt wie bei unseren übrigen Leistungen zu den unter Was es kostet beschriebenen Bedingungen.

Wenn Agenten bereits Schlüssel zu Ihren Systemen besitzen,
sagen Sie uns, welche.

Jeder Agent, den Sie betreiben, **hat einen Schlüssel zu etwas** — wissen Sie, welchen und wozu.

Es beginnt rein lesend: mit einer Übersicht der einbezogenen Agenten und Zugangsdaten sowie nach Risiko geordneten Befunden.

Eine Agentenübersicht anfragen Was es kostet →

Den Umfang vor der Beauftragung festlegen

Beschreiben Sie den Bereitstellungsweg, die Betriebsrisiken und die Bereitschaftsverantwortlichen. Wir können eine Änderung mit Prüfungen, einem Wiederherstellungsweg und klaren Zuständigkeiten abgrenzen.

Preise ansehen Kontakt aufnehmen →

Ihre Anforderungen besprechen

Erzählen Sie uns, was Sie von „Sicherheit für Ihre KI-Agenten“ benötigen, einschließlich des Termins und des Teams, das die Wartung übernehmen wird. Wir können einen praktikablen Weg besprechen.

Unsere Leistungen entdecken

welcome@engineer.company +45 50 10 61 84 Kopenhagen, Dänemark

Nehmen Sie diese Seite mit

Kontaktkarte

Mit Sorgfalt und Liebe gestaltet von D.

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

Copyright © 2025 – heute · Engineer ApS · Alle Rechte vorbehalten.

Gedruckt von https://ditte.engineer.company/de/ai-security/