Engineer

Sikkerhed for AI-agenter i din virksomhed

Gennemgang af AI-agenters adgang: aftalt kortlægning,
prioriterede fund, rettighedstest og dokumentation for kontrollen.

Få overblik over, hvad jeres AI-agenter kan nå, og styr adgangen. En agent med forbindelse til mail, dokumenter eller et API giver rettigheder og fejlmuligheder, der skal gennemgås. Vi kortlægger de aftalte systemer, vurderer risici, indfører godkendte sikkerhedsforanstaltninger og afprøver de begrænsede handlinger. I får adgangskortet, fundene og dokumentationen for kontrollen.

Gør ydelsen til et resultat

Lad »Sikkerhed for AI-agenter i din virksomhed« understøtte et forretningsmål. Vi forbinder det tekniske arbejde med de brugere, den drift og de beslutninger, det skal støtte. Hver agents adgang kortlagt og begrænset til det nødvendige, hver handling logget, hvor agenten ikke kan ændre loggen, og hver kontrol afprøvet frem for taget for givet.

Se vores ydelser Kontakt os →

Undersøg det relevante arbejde

Brug arbejdet på siden til at undersøge vores tilgang. Spørg til afvejningerne og forudsætningerne bag hvert resultat.

Cases Kontakt os →

Kortlagt: hver agent, og hvad den kan nå

Vi kortlægger agenter inden for det aftalte omfang, både godkendte integrationer og andre forbindelser, som gennemgangen med jeres medarbejdere viser. Kortet beskriver legitimationsoplysninger, værktøjer og tilgængelige data for hver agent. I får kortet sammen med fundene, og vi aftaler ansvaret for opdatering ved overdragelsen. Gennemgangens omfang og huller fremgår udtrykkeligt.

Låst: hver agent med mindste privilegium

En agent får sin egen identitet og kun den adgang, dens opgave kræver — mindste privilegium: roller i stedet for delte konti (RBAC), kun læseadgang som udgangspunkt og skriveadgang kun gennem snævert afgrænsede værktøjer med loft over, hvor ofte og hvor meget. Hemmeligheder deles op pr. agent og pr. vært, indsættes aldrig i en prompt og roteres uden en kodeændring. Ekstern tekst, en agent læser — en e-mail, en webside eller et dokument — behandles som data, der ikke er tillid til. Værktøjernes rettigheder og valideringskontroller håndhæver de aftalte grænser; en besked med »ignorer dine regler« er ikke en tilladelse til at udføre en handling. Kontrollerne er fail-closed: kan et tjek ikke køre, bliver handlingen afvist.

Registreret: hvad den gjorde, og for hvem

Hvert værktøjskald logges med den person, det blev foretaget på vegne af, hvad agenten blev bedt om, hvad den rørte ved, og hvad der kom tilbage, og loggen gemmes et sted, hvor agenten ikke selv kan ændre den. Alarmregler aftales for systemet. De kan omfatte en pludselig stigning i antallet af læsninger, brug uden for de forventede tider eller adgang til data uden for agentens aftalte dataområde. Går noget alligevel galt, er det loggen, der gør en sikkerhedshændelse til en tidslinje i stedet for et gæt.

Et kort, en lås og en log **er tre påstande** —
indtil nogen prøver at komme forbi dem.

De aftalte kontroller afprøves med repræsentative adgangs-, misbrugs- og gendannelsesscenarier før overdragelsen.

En lås skal afprøves, ikke tages for givet

Vi tester adgangskontroller inden for opgavens omfang og dokumenterer, hvad der blev forsøgt, afvist eller kræver mere arbejde. Test kan omfatte afviste værktøjskald, ratebegrænsning, adskilte legitimationsoplysninger og forsøg på prompt-injektion. Resultatet er dokumentation og gentagelige

kontroller med beskrevne grænser. Hændelseshåndtering og rotation af legitimationsoplysninger aftales særskilt, når et eksisterende læk skal begrænses.

Arbejdet, der ligger bag

Vi har udført dette arbejde på servere, i identitetssystemer og på platforme. Vi skrev 1.000 RBAC-regler til Linux- og databaseservere, ledede adgangsstyringen i M365 og single sign-on og kørte et løbende sikkerhedsscanningsprogram med 21 skriftlige revisioner. Vi adskilte alle driftshemmeligheder pr. vært, tilføjede et preflight-tjek af hemmeligheder på ét sekund, fandt tre låse på døren, og ingen af dem var låst, og beviste, at indtrængningsblokeringen virker ved hver kørsel. Vi byggede fail-closed rate limiting mod misbrug og skrev en afgrænsningsregel efter et læk. Hvordan vi selv holder orden i eget hus, kan du læse under sådan arbejder vi sikkert.

Sådan forløber et samarbejde

Det begynder med ren læseadgang. Vi kortlægger agenterne inden for det aftalte omfang, deres adgangsoplysninger og det, de giver adgang til, og giver dig fundene sorteret efter risiko, hvert med sin begrundelse og sin rettelse. Derefter laves rettelserne i din konfiguration og dit repository, gennem dit review og med tjek, der slår fejl, når de kontrollerede adgangsveje giver adgang ud over de aftalte rettigheder. Du beholder kortet, loggen og testene — den mappe, du rækker en revisor eller et forsikringsselskab — og de data, dine agenter læser, kan adgangsstyres på samme måde. Det købes som resten af vores arbejde, på vilkårene i hvad det koster.

Har agenter allerede nøgler til dine systemer, så fortæl os, hvilke.

Hver agent, I bruger, **har en nøgle til noget** — find ud af hvilken, og til hvad.

Det begynder med ren læseadgang: et kort over de agenter og adgangsoplysninger, opgaven omfatter, med fundene sorteret efter risiko.

Få dine agenter kortlagt Hvad det koster →

Afgræns opgaven før bestilling

Medbring udrulningsforløbet, driftsrisiciene og dem, der har vagten. Vi kan afgrænse en ændring med kontroller, en vej tilbage og tydeligt ansvar.

Se, hvad det koster Kontakt os →

Gennemgå jeres krav med os

Fortæl os, hvad I har brug for inden for »Sikkerhed for AI-agenter i din virksomhed«, herunder fristen og det team, der skal vedligeholde løsningen. Vi kan drøfte en praktisk vej frem.

Se vores ydelser

welcome@engineer.company +45 50 10 61 84 København, Danmark

Tag denne side med dig

Visitkort

Lavet med kærlighed af D.

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

Copyright © 2025 – nu · Engineer ApS · Alle rettigheder forbeholdes.

Udskrevet fra https://daryna.engineer.company/da/ai-security/