Engineer

Sikkerhet for AI-agentene du kjører

Gjennomgang av AI-agenters tilgang: avtalt kartlegging, prioriterte funn, rettighetstester og dokumentasjon fra kontrollen.

Få oversikt over hva AI-agentene deres når, og sett kontroll på tilgangen. En agent koblet til e-post, dokumenter eller et API gir rettigheter og feilveier som må gjennomgås. Vi kartlegger de avtalte systemene, vurderer risiko, innfører godkjente sikkerhetstiltak og prøver de begrensede handlingene. Dere får tilgangskartet, funnene og dokumentasjonen fra kontrollen.

Gjør tjenesten til et resultat

La «Sikkerhet for AI-agentene du kjører» støtte et forretningsmål. Vi kan knytte det tekniske arbeidet til brukerne, driften og beslutningene det skal støtte. Hver agents tilgang kartlagt og begrenset til det nødvendige, hver handling logget der agenten ikke kan endre loggen, og hvert sikkerhetstiltak testet i stedet for tatt på tro.

Se tjenestene våre Ta kontakt →

Undersøk det relevante arbeidet

Bruk arbeidet på denne siden til å undersøke tilnærmingen vår. Spør om avveiningene og forutsetningene bak hvert resultat.

Prosjekter Ta kontakt →

Kartlagt: hver agent og hva den når

Vi kartlegger agenter innenfor avtalt omfang, både godkjente integrasjoner og andre forbindelser gjennomgangen med de ansatte avdekker. Kartet viser påloggingsopplysninger, verktøy og tilgjengelige data for hver agent. Dere får kartet sammen med funnene, og vi avtaler ansvar for oppdatering ved overlevering. Gjennomgangens omfang og hull er uttrykkelig beskrevet.

Låst: hver agent med minste privilegium

Hver agent får sin egen identitet og ikke flere privilegier enn jobben krever: roller i stedet for delte kontoer (RBAC), kun lesetilgang som standard, og skrivetilgang bare gjennom avgrensede verktøy med grenser for hvor ofte og hvor mye. Hemmeligheter deles opp per agent og per vert, limes aldri inn i en prompt og kan roteres uten kodeendringer. Ekstern tekst en agent leser — en e-post, en nettside eller et dokument — behandles som data man ikke har tillit til. Verktøyenes rettigheter og valideringskontroller håndhever de avtalte grensene; en melding med «ignorer reglene dine» er ikke tillatelse til å utføre en handling. Kontrollene feiler lukket: når en sjekk ikke kan kjøres, avvises handlingen.

Loggført: hva den gjorde, og for hvem

Hvert verktøykall logges med hvem det ble gjort for, hva som ble bedt om, hva det berørte og hva som kom tilbake, og loggen oppbevares der agenten selv ikke kan endre den. Varslingsregler avtales for systemet. De kan omfatte en plutselig økning i antall lesinger, bruk utenfor forventede tider eller tilgang til data utenfor agentens avtalte dataområde. Skjer noe likevel, er det loggen som gjør en hendelse til en tidslinje i stedet for gjetting.

Et kart, en lås og en logg **er tre påstander** —
helt til noen prøver å komme seg forbi dem.

De avtalte kontrollene prøves med representative tilgangs-, misbruks- og gjenopprettingsscenarioer før overleveringen.

En lås skal testes, ikke tas på tro

Vi tester tilgangskontroller innenfor oppdragets omfang og dokumenterer hva som ble forsøkt, avvist eller trenger mer arbeid. Tester kan omfatte avviste verktøykall, ratebegrensning, isolerte påloggingsopplysninger og forsøk på prompt-injeksjon. Resultatet er dokumentasjon og gjentakbare kontroller med beskrevne grenser. Hendelsesarbeid og rotasjon av påloggingsopplysninger avtales separat når en eksisterende lekkasje må begrenses.

Arbeidet bak det

Vi har gjort dette arbeidet på servere, identitetssystemer og plattformer. Vi skrev 1 000 RBAC-regler på tvers av Linux- og databaseservere, ledet tilgangsstyringen på tvers av M365 og single sign-on, og drev et kontinuerlig program for sikkerhetsskanning med 21 skriftlige revisjoner. Vi delte opp hver driftshemmelighet per vert, la til en forhåndssjekk av hemmeligheter på ett sekund, fant tre låser på døren, og ingen av dem var låst, og beviste at utestengingen av inntrengere virker ved hver kjøring. Vi bygde hastighetsbegrensning som feiler lukket og skrev en avgrensningsregel etter en lekkasje. Hvordan vi holder orden i eget hus, står på slik arbeider vi sikkert.

Slik foregår et oppdrag

Det begynner med en gjennomgang med kun lesetilgang. Vi kartlegger agentene innenfor avtalt omfang, påloggingsopplysningene deres og hva de gir tilgang til, og gir deg funnene sortert etter risiko, hvert med begrunnelse og forslag til løsning. Rettelsene gjøres deretter i konfigurasjonen og repoet ditt og går gjennom din egen kodegjennomgang, med sjekker som feiler når de kontrollerte tilgangsveiene gir tilgang utover de avtalte rettighetene. Du beholder kartet, loggen og testene — det du gir en revisor eller et forsikringsselskap — og dataene agentene leser, kan tilgangsstyres på samme måte. Det kjøpes som resten av arbeidet vårt, på vilkårene i hva det koster.

Har agenter allerede nøkler til systemene dine, så si hvilke.

Hver agent dere kjører, **har en nøkkel til noe** — og du bør vite hvilken, og til hva.

Det begynner med kun lesetilgang: et kart over agentene og tilgangsopplysningene oppdraget omfatter, med funnene sortert etter risiko.

Be om et agentkart Hva det koster →

Avgrens oppgaven før bestilling

Ta med utrullingsløpet, driftsrisikoene og dem som har vakt. Vi kan avgrense en endring med kontroller, en vei tilbake og tydelig ansvar.

Se hva det koster Ta kontakt →

Gå gjennom kravene med oss

Fortell hva dere trenger innen «Sikkerhet for AI-agentene du kjører», inkludert fristen og teamet som skal vedlikeholde løsningen. Vi kan drøfte en praktisk vei videre.

Se tjenestene våre

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

Ta med deg denne siden

Visittkort

Laget med kjærlighet av D.

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

Opphavsrett © 2025 – nå · Engineer ApS · Alle rettigheter forbeholdt.

Skrevet ut fra https://ditte.engineer.company/nb/ai-security/