Engineer Mjukvara

IT-säkerhet för de AI-agenter du kör

Granskning av AI-agenters åtkomst: avtalad kartläggning, prioriterade resultat, behörighetstester och dokumentation från kontrollen.

Få överblick över vad era AI-agenter når och kontrollera åtkomsten. En agent ansluten till e-post, dokument eller ett API ger behörigheter och möjliga felvägar som behöver granskas. Vi kartlägger avtalade system, bedömer risker, inför godkända säkerhetsåtgärder och prövar begränsade handlingar. Ni får åtkomstkartan, resultaten och dokumentationen från kontrollen.

Gör tjänsten till ett resultat

Låt ”IT-säkerhet för de AI-agenter du kör” stödja ett affärsmål. Vi kan knyta det tekniska arbetet till användarna, driften och besluten det ska stödja. Varje agents behörigheter kartlagda och begränsade till det nödvändiga, varje handling loggad där agenten inte kan ändra loggen, och varje skydd prövat i stället för taget för givet.

Se våra tjänster Hör av er →

Granska det relevanta arbetet

Använd arbetet på sidan för att granska vårt angreppssätt. Fråga om avvägningarna och förutsättningarna bakom varje resultat. Byggde hastighetsbegränsning som stänger vid fel — 22 hastighetsbegränsare som stänger vid fel

Byggde hastighetsbegränsning som stänger vid fel Hör av er →

Kartlagd: varje agent och vad den når

Vi kartlägger agenter inom avtalad omfattning, både godkända integrationer och andra anslutningar som genomgången med era medarbetare visar. Kartan beskriver inloggningsuppgifter, verktyg och åtkomliga data för varje agent. Ni får kartan med resultaten, och ansvaret för uppdatering avtalas vid överlämning. Granskningens omfattning och luckor beskrivs uttryckligen.

Låst: varje agent på minsta behörighet

En agent får en egen identitet och den minsta behörighet dess uppgift kräver: roller i stället för delade konton (RBAC), skrivskyddad som standard, skrivningar bara genom snäva verktyg med gränser för hur ofta och hur mycket. Hemligheter delas upp per agent och per värd, klistras aldrig in i en prompt och byts ut utan en kodändring. Extern text som en agent läser — ett e-postmeddelande, en webbsida eller ett dokument — behandlas som data som inte är betrodda. Verktygens behörigheter och valideringskontroller upprätthåller de avtalade gränserna; ett meddelande med ”ignorera dina regler” ger inte tillstånd att utföra en handling. Kontrollerna stänger vid fel: när en kontroll inte kan köras nekas handlingen.

Loggad: vad den gjorde, och åt vem

Varje verktygsanrop loggas med personen det gjordes åt, vad agenten ombads göra, vad den rörde och vad som kom tillbaka, och loggen förvaras där agenten själv inte kan ändra den. Larmregler avtalas för systemet. De kan omfatta en plötslig ökning av antalet läsningar, användning utanför förväntade tider eller åtkomst till data utanför agentens avtalade dataområde. När något väl går fel är det loggen som gör en incident till en tidslinje i stället för en gissning.

En karta, ett lås och en logg **är tre påståenden** —
tills någon försöker ta sig förbi dem.

De avtalade kontrollerna prövas med representativa scenarier för åtkomst, missbruk och återställning före överlämningen.

Ett lås prövas, det tas inte på tro

Vi testar åtkomstkontroller inom uppdragets omfattning och dokumenterar vad som prövades, avvisades eller behöver mer arbete. Tester kan omfatta nekade verktygsanrop, hastighetsgränser, isolerade inloggningsuppgifter och försök till promptinjektion. Resultatet är dokumentation och upprepningsbara kontroller med beskrivna gränser. Incidentarbete och byte av inloggningsuppgifter avtalas separat när en befintlig exponering behöver begränsas.

Arbetet bakom

Vi har gjort det här arbetet på servrar, identitetssystem och plattformar. Vi skrev 1 000 RBAC-regler för Linux- och databasservrar, ledde behörighetshanteringen i M365 och för single sign-on och drev ett löpande program för säkerhetsskanning med 21 skriftliga granskningar. Vi delade upp varje driftshemlighet per värd, lade till en förhandskontroll av hemligheter som tar en sekund, hittade tre lås på dörren, inget av dem låst, och bevisade att intrångsspärren fungerar vid varje körning. Vi byggde hastighetsbegränsning som stänger vid fel och skrev en avgränsningsregel efter en läcka. Hur vi sköter vårt eget hus står på så arbetar vi säkert.

Så går ett uppdrag till

Det börjar med enbart läsbehörighet. Vi kartlägger agenterna inom avtalad omfattning, deras inloggningsuppgifter och vad de ger åtkomst till, och lämnar över fynden i riskordning, vart och ett med sitt skäl och sin åtgärd. Åtgärderna görs sedan i din konfiguration och ditt repo, genom din granskning, med kontroller som misslyckas när de kontrollerade åtkomstvägarna ger åtkomst utöver de avtalade behörigheterna. Du behåller kartan, loggen och testerna — den fil du lämnar till en revisor eller ett försäkringsbolag — och datan agenterna läser kan behörighetsstyras på samma sätt. Det köps som resten av vårt arbete, på villkoren i vad det kostar.

Om agenter redan har nycklar till dina system, berätta vilka.

Varje agent ni kör **har en nyckel till något** — ta reda på vilken, och till vad.

Det börjar med enbart läsbehörighet: en karta över agenterna och inloggningsuppgifterna som ingår i uppdraget, med fynden i riskordning.

Be om en agentkarta Vad det kostar →

Avgränsa uppdraget före beställning

Ta med driftsättningsvägen, driftsriskerna och dem som har jour. Vi kan avgränsa en förändring med kontroller, en väg tillbaka och tydligt ansvar.

Se vad det kostar Hör av er →

Gå igenom era krav med oss

Berätta vad ni behöver inom ”IT-säkerhet för de AI-agenter du kör”, vilken tidsram ni önskar och vilket team som ska förvalta lösningen. Vi kan diskutera en praktisk väg framåt.

Se våra tjänster

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

Ta med dig den här sidan

Visitkort

Omsorgsfullt skapad med kärlek av D.

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

Upphovsrätt © 2025 – nu · Engineer ApS · Alla rättigheter förbehållna.

Utskrivet från https://software.engineer.company/sv/ai-security/