ilionx
experts in eenvoud

editie najaar 2026

laat je bestuur zien welke agents er draaien en van wie ze zijn

Preplijst voor CISO's: 6 voorwaarden voordat een agent in je systemen mag

Je weet ongeveer wat er draait, maar dit aantonen is andere koek. De business bouwt agents in platforms waar je al een licentie op hebt. Ze werken, ze zijn nuttig en niemand meldt ze proactief bij jou aan. Jij hoort ervan als er iets misgaat of als je er zelf naar vraagt.

Sinds 15 augustus 2026 is je bestuur wettelijk aanspreekbaar op cyberrisico. Ze willen niet weten wat er ongeveer draait. Ze willen bewijzen dat het goed zit.

Om die vraag zo goed mogelijk te beantwoorden heb je iets anders nodig: een agent-register met je belangrijkste agents erin.

wat je hieruit haalt

- **Deel 1** is de onderbouwing: zes voorwaarden waaraan een agent voldoet voordat hij in je systemen mag en wie er per voorwaarde aan zet is.
- **Deel 2** levert een register van je tien belangrijkste agents: wie de eigenaar is, wat ze mogen en waar je ze kunt tegenhouden.

waar moet je starten

Als dit voor jou klopt	Begin bij
Je moet binnenkort met je bestuur praten over AI-agents en je weet er zelf nog weinig van	deel 1
Je weet niet precies hoeveel agents er draaien of wie de eigenaren zijn	deel 2

De meeste organisaties weten niet hoeveel agents er draaien. Bij hen ging het bouwen sneller dan het goedkeuren. Van de organisaties die aan AI-agents begonnen zijn, ging 81% al van plan naar test of productie. Bij 14% van hen heeft security akkoord gegeven op alle agents die draaien *(bron: Gravitee, State of AI Agent Security 2026)*.

de business bouwt, security keurt goed. niet in hetzelfde tempo

81% van de teams zit voorbij de planfase, in test of productie

14% heeft securitygoedkeuring op alle agents die draaien

bron: Gravitee, State of AI Agent Security 2026 (februari 2026), 900+ respondenten. Twee losse cijfers uit dezelfde meting.

Deel 1
grip krijgen op AI-agents

wat er is veranderd

Tot vorig jaar beoordeelde je AI op wat de output was, een model gaf antwoord, iemand keek of het klopte en het risico zat in de inhoud van dat antwoord.

Dat geldt nog steeds, maar er is ook iets bijgekomen, een agent geeft ook output, maar neemt vooral actie. Hij leest een mailbox, schrijft in een systeem, benadert een API, stuurt een bericht naar buiten. Heeft rechten. En werkt door als niemand kijkt.

Daarmee verschuift de vraag van “klopt de output” naar “wat heeft de agent gedaan en mocht dat”. Dat is een andere vraag en de meeste organisaties hebben er nog geen antwoord op.

Drie feiten die de urgentie bepalen:

- **Je bestuur is aanspreekbaar:** Sinds 15 augustus 2026 geldt de Cyberbeveiligingswet voor ongeveer 8.000 Nederlandse organisaties. Het bestuur is eindverantwoordelijk voor de beheersing van cyberrisico.
- **Je moet kunnen aantonen wat je AI doet:** Sinds 2 augustus 2026 gelden de deployer-verplichtingen uit de EU AI Act: menselijk toezicht en logbewaring.
- **De aanvaller is sneller geworden:** De gemiddelde breakout time bij criminele aanvallen lag in 2025 op 29 minuten, 65% korter dan het jaar ervoor. De snelst waargenomen breakout was 27 seconden. AI-ondersteunde operaties namen met 89% toe.

twee data in augustus 2026

2 augustus 2026
EU AI Act, deployer-verplichtingen van kracht, waaronder menselijk toezicht en logbewaring (artikel 26).

15 augustus 2026
Cyberbeveiligingswet en Wet weerbaarheid kritieke entiteiten van kracht. Bestuur keurt maatregelen goed en houdt toezicht, circa 8.000 organisaties.

meldplicht na een incident

24 uur
eerste melding
bij het CSIRT

72 uur
nadere melding,
met eerste beoordeling

1 maand
eindverslag naar
CSIRT en toezichthouder

bron: Rijksoverheid en NCSC; artikel 26 EU AI Act

wat een agent is

Een agent is software die een taak van begin tot eind uitvoert zonder dat iemand per stap meekijkt. Wie daarbij aan ChatGPT of Claude denkt, ziet maar een klein deel van wat er in je organisatie draait. Je vindt agents op drie plekken.

SAAS-PLATFORM

agents in software die je al gebruikt

De business bouwt agents in een platform waar je al een licentie op hebt. Groeit het snelst, en hier kom je het lastigst bij de logs.

BIJVOORBEELD

- Copilot Studio
- ServiceNow
- Salesforce
- Azure Foundry

EIGEN CLOUD

agents die je zelf bouwt

Je eigen team bouwt agents op je eigen cloud. Volledige controle over de harness, vaak zonder stopknop.

LOKAAL

agents op de laptop van je medewerkers

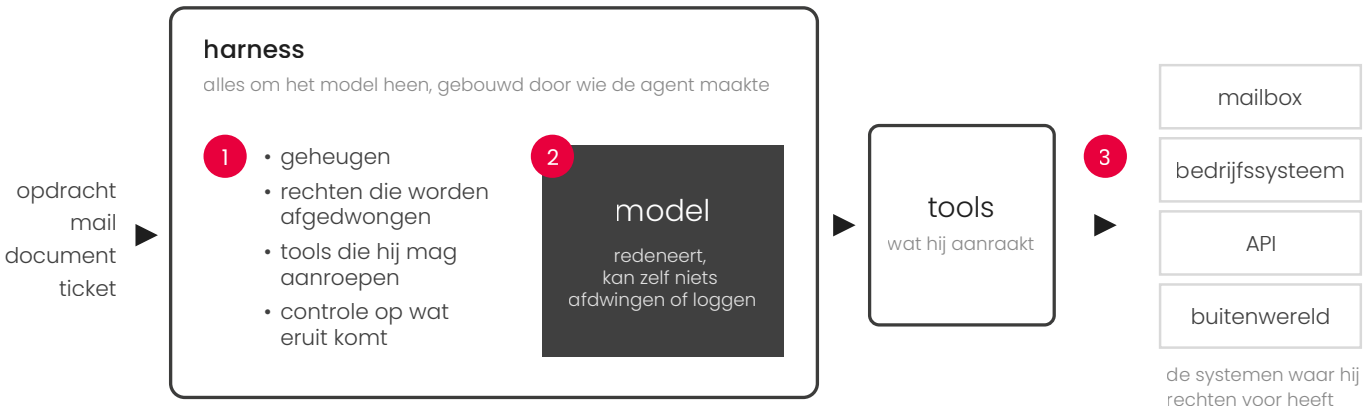
Een medewerker draait een agent lokaal, gekoppeld aan bedrijfssystemen. Dit zie je vooral bij ontwikkelaars.

BIJVOORBEELD

- Claude
- Codex
- Cursor

model plus harness

Een agent is een model plus een harness. Het model redeneert, de harness regelt de rest: geheugen, tools, rechten en de controle op wat eruit komt. Het model kan zelf geen rechten afdwingen en niets loggen. Wie zegt "we gebruiken een veilig model", zegt dus weinig over of de agent beheerst is.



1 in de harness
beperken wat de agent überhaupt mag proberen
beheer: degene die de agent bouwde

2 bij het model
zien en filteren wat er in en uit gaat
beheer: je eigen team, als je een gateway hebt

3 bij de tools
blokkeren wat de agent daadwerkelijk aanraakt
beheer: de beheerder van dat systeem

Welke van de drie plekken je zelf in handen hebt, hangt af van de soort agent. Bij andermans SaaS zijn het vaak alleen de tools, bij zelfbouw alle drie. Bij een lokale agent zie je meestal alleen het netwerk.

ilionx

het uitgangspunt

Beheersing begint bij twee dingen: weten wat er draait en per agent een plek hebben waar zijn handelingen worden gezien en zo nodig geweigerd. Die plek is pas iets waard als wat je daar ziet nog dezelfde dag bij iemand terechtkomt. Logs die je pas na een incident leest, zijn voor de meldplicht van 24 uur te laat. Achteraf moet je bovendien kunnen laten zien wat er is gebeurd. De meeste organisaties regelen het weten, kopen een platform voor het meekijken en komen aan het aantonen niet toe. Bij de toezichthouder gaat het juist op dat laatste mis.

waar dit op aansluit

Onze aanpak voor veilige AI-agents kent vier stappen voor de organisatie: weten wat er draait, kunnen meekijken, kunnen ingrijpen en kunnen aantonen. Per agent worden dat zes voorwaarden.

Stap voor de organisatie	Voorwaarde per agent	ISO 42001
1. weten wat er draait	1. herkomst 2. identiteit	A.4.2, A.4.4, A.9.4 identiteit heeft geen eigen control
2. kunnen meekijken	3. onderschepping 4. waarneming	A.6.2.5, A.6.2.6
3. kunnen ingrijpen	5. ingrijpen	A.6.2.6, A.8.4
4. kunnen aantonen	6. verantwoording	A.6.2.7, A.6.2.8

de zes voorwaarden voordat een AI-agent in je systemen mag

1 herkomst
de agent staat geregistreerd, met een eigenaar en een doel

Je weet dat hij bestaat, wie hem bouwde, wat hij hoort te doen en welke persoon erover gaat, met telefoonnummer. Het is de saaiste voorwaarde en ook de vaakst overgeslagen, nu iedereen met een licentie zelf agents bouwt.

Zonder deze voorwaarde: je kunt bij een incident niemand bellen en je weet niet of afwijkend gedrag afwijkend is.

2 identiteit
de agent is iemand met eigen rechten

Een eigen account op least privilege, nooit een gedeeld serviceaccount of de inlog van een medewerker. Hier zit de grootste blinde vlek: 69% deelt ergens inloggegevens tussen agents en maar 22% geeft agents een eigen identiteit.

Zonder deze voorwaarde: je kunt de schade niet afbakenen. Deze ene agent uitzetten kan ook niet zonder alles mee te trekken wat op dezelfde identiteit draait.

3 onderschepping
er is een plek waar handelingen van een agent langs een controle gaan

Dit is een architectuurkeuze: in de harness, bij het model of bij de tools. Welke plek past, hangt af van waar de agent draait. Meestal heb je er meer dan één nodig.

Zonder deze voorwaarde: je beleid staat op papier, maar leeft niet in de praktijk.

4 waarneming
gedrag komt binnen bij iemand die meekijkt

Gedrag van een agent komt binnen bij een team dat 24 uur per dag meekijkt, in plaats van op een dashboard dat je eens per week opent. Omdat een agent meestal één vaste taak heeft, valt afwijkend gedrag snel op. Toch wordt gemiddeld maar 47% van de agents gemonitord of gelogd.

Zonder deze voorwaarde: je detectie hangt af van iemand die belt.

5 ingrijpen
je kunt de agent stoppen en je weet wie dat mag

Je kunt één agent binnen minuten pauzeren of intrekken zonder een bedrijfsproces plat te leggen. Er is een persoon die dat besluit mag nemen, ook 's nachts. Eenduidige stappen zoals een sessie intrekken of isoleren mogen automatisch, bij beslissingen met impact zet je een mens.

Zonder deze voorwaarde: je kijkt naar het incident terwijl het doorloopt.

6 verantwoording
je kunt achteraf aantonen wat een agent deed

Een audit trail per agent, logs die lang genoeg bewaard blijven en een rapportage die je bestuur en toezichthouder herkennen. Voor hoogrisico-systemen vraagt de AI Act minimaal zes maanden. Let bij een incident op de volgorde, want veel SaaS-platforms overschrijven de logs zodra je een account intrekt.

Zonder deze voorwaarde: je hebt het incident opgelost, maar kunt het niet uitleggen.

de zes voorwaarden hangen van elkaar af

Vier van de zes is geen voldoende. Zonder waarneming lees je je onderscheppingslogs pas na afloop. Zonder verantwoording kun je een opgelost incident niet uitleggen aan de toezichthouder.

ilionx

wie doet wat

Wees duidelijk over de rolverdeling, want daar zit vaak de verwarring.

Voorwaarde 1, 2 en 3 zijn architectuur. Een platform zoals Noma vindt alle agents, ook de shadow-agents, en zet rechten op least privilege. Het laat ook gevaarlijke combinaties zien, zoals een agent die code uitvoert en ook naar buiten praat.

Voorwaarde 4 en 5 zijn operatie. Iemand moet weten wat normaal gedrag is en een draaiboek klaar hebben als een agent afwijkt. Dat doen wij met onze partners vanuit ons Nederlandse SOC.

Voorwaarde 6 is van jou en je bestuur. Wij leveren inzage en rapportage,hetgesprekvoerjezelf. Je kunt dit ook met een andere partij doen. Wat niet werkt, is de architectuur kopen en ervan uitgaan dat de operatie vanzelf volgt.

wat je hier invult, gebruik je aan de bestuurstafel

Je bestuur stelt het komend jaar steeds meer vragen over AI-security. De belangrijkste hebben we uitgewerkt in *De 12 vragen die je RvB over AI-security gaat stellen*. Per vraag vind je daar het antwoord dat je klaar wilt hebben.

Dit document levert het materiaal voor zeven van de twaalf vragen. Je bent dus al een heel eind op weg.

Wat je hier invult	Welke bestuursvraag dat beantwoordt
Aantal agents en hun verdeling over de drie soorten	“Hoeveel AI draait er nu eigenlijk bij ons?”
Rechten per agent en welke van de vijf eigenschappen hij heeft	“Kan zo’n agent bij ons echt schade aanrichten?”
Wie de noodstop mag gebruiken en hoe dat gaat	“Kunnen we AI niet gewoon verbieden tot het veilig is?”
Per SaaS-agent: kun je hem apart stoppen en kom je bij de logs?	“En wat als het bij een van onze leveranciers misgaat?”
Logging per agent en de bewaartermijn	“Als die AI bij ons een besluit neemt, kunnen we dan uitleggen waarom?”
Per agent een eigenaar met naam en de datum waarop je deze het laatst controleerde	“En ben ik nou wel of niet persoonlijk aansprakelijk?”
Je zwakste schakel en de eerste stap daarop	“Goed, ik snap het belang. Wat doen we morgen anders?”

> ‘je kunt geen agent beheersen waarvan je niet weet dat hij bestaat’

Deel 2
de preplijst

wat je op papier moet hebben voordat je ergens over kunt beslissen

Deze lijst is voor de situatie waarin de meeste organisaties nu zitten, er draaien agents, het overzicht ontbreekt en je bestuur wil met je in gesprek. Je vult de lijst zelf in, met je eigen team. Dat kost je een paar uur als je de antwoorden al op een rijtje hebt en iets langer als je informatie nog moet ophalen in de organisatie.

begin niet bij alle agents

Een volledige inventarisatie klinkt als de juiste eerste stap. In de praktijk duurt die maanden en is de lijst verouderd voordat hij af is. Loop je agents langs en stel per agent drie vragen:

- Heeft deze agent schrijfrechten in een systeem dat ertoe doet? (finance, HR, klantdata, productie)
- Praat hij met de buitenwereld? (mail, klantportaal, leveranciers, publieke API)
- Is hij gebouwd buiten IT om? (door de business, in een SaaS-platform, door een individuele developer)

Twee keer ja of vaker, die agent gaat op je lijst.

Kom je boven de tien uit, begin dan bij degene met de meeste rechten. Kun je deze drie vragen niet beantwoorden omdat je geen overzicht hebt van wat er draait? Dan is dat je eerste bevinding en die hoort in je bestuursgesprek thuis.

waar een agent draait, bepaalt wat je ziet

In deel 1 staan de drie soorten en waar je per soort kunt ingrijpen. Hier staat per soort wat je al hebt, wat je waarschijnlijk mist en welke vraag je moet stellen voordat je de agentkaart invult.

In andermans SaaS: De business bouwde de agent in een platform waar de organisatie een licentie op heeft, zoals Copilot Studio, Salesforce, ServiceNow of Azure Foundry.

- Wat je hebt: wat de leverancier je geeft.
- Wat je mist: vaak de harness zelf. Je kunt niet altijd bij de logs en niet altijd één agent apart uitzetten.
- Vraag: kan ik in dit platform één agent stoppen zonder het account van het hele team te raken en hoe lang bewaart de leverancier de logs?

Zelfgebouwd: Je eigen team bouwde de agent op je eigen cloud.

- Wat je hebt: volledige controle over de harness. Dit is de enige soort waar je alle drie de onderscheppingspunten zelf kunt inrichten.
- Wat je mist: meestal een logregel per handeling, een stopknop en een controle die een handeling kan weigeren.
- Vraag: zit er in de code een plek waar het systeem een handeling kan weigeren en wie mag dat aanpassen?

Lokaal: Een medewerker draait de agent op zijn laptop, met Claude, Cursor of Codex, gekoppeld aan bedrijfssystemen. Vooral ontwikkelaars.

- Wat je hebt: endpointbeheer en zicht op het netwerk.
- Wat je mist: je ziet zelden welke tools de agent aanroept en de rechten zijn die van de medewerker zelf.
- Vraag: welke koppelingen heeft deze agent naar systemen die niet van deze medewerker zijn?

ilionx

de agentkaart

Per agent één kaart, reken op een uur als je veel informatie al voorhanden hebt en een week als je alles nog moet uitzoeken.

veld	antwoord	voorwaarde
Naam en doel van de agent		1
Eigenaar: naam en telefoonnummer		1
Gebouwd door, sinds wanneer		1
Soort: lokaal / zelfgebouwd / andermans SaaS		3
Praat hij, doet hij, of allebei		3
Identiteit: eigen of gedeeld — welk account		2
Leesrechten: welke systemen en data		2
Schrijfrechten: welke systemen		2
Tools die hij aanroept en van wie die zijn		2
Onderscheppingspunt: harness / model / tools		3
Wie beheert dat onderscheppingspunt		3
Wie ziet wat deze agent doet en wanneer kijkt die persoon		4
Noodstop: wie mag het, hoe, in hoeveel minuten		5
Wat gaat er stuk als je hem stopt		5
Logging: waar en hoe lang bewaard		6
Laatst gecontroleerd op		1
Kun je binnen 24 uur reconstrueren wat deze agent deed?	☐ ja ☐ nee ☐ niet getest	6

Let vooral op het onderscheid tussen praten en doen. Een agent die alleen antwoord geeft, is een inhoudelijk risico. Een agent die handelingen uitvoert, is een operationeel risico. Ze vragen andere maatregelen en in de praktijk halen mensen ze door elkaar.

wat een agent kan en waarom de combinatie het probleem is

Wat een agent nuttig maakt, maakt hem ook risicovol. De Cloud Security Alliance benoemt vijf eigenschappen die dat bepalen.

Kruis per agent aan wat van toepassing is:

- ☐ Komt bij gevoelige data: Klantgegevens, contracten, personeelsdossiers.
- ☐ Praat naar buiten: Met het open internet, of met partijen buiten je organisatie.
- ☐ Beweegt tussen systemen: Zelfstandig, van de ene omgeving naar de andere.
- ☐ Verwerkt tekst van buiten: Mail, documenten, webpagina's, tickets.
- ☐ Handelt zonder mens: Hij voert uit zonder dat iemand ertussen zit.

Het risico zit in de combinatie. Vier combinaties waar je naar moet kijken:

Combinatie	Wat er kan gebeuren
Onvertrouwde invoer + autonoom handelen	De agent voert een instructie uit die in een binnenkomend document stond. Dit is de meest onderschatte route.
Diepe datatoegang + externe connectiviteit	Interne gegevens verlaten je omgeving zonder dat er iets is gehackt.
Laterale beweegruimte + gedeelde identiteit	Je kunt de schade niet afbakenen, want je weet niet waar de identiteit ophoudt.
Autonoom handelen + geen logging	Je lost het incident op en kunt daarna niet uitleggen wat er gebeurde.

Drie of meer vinkjes: deze agent heeft je aandacht nodig.

wat je hierna hebt

Als je dit voor tien agents hebt ingevuld, heb je:

een register

Niet compleet, wel je belangrijkste agents op een rij. Je laat zien dat je de grootste risico's kent, en dat bedoelt de zorgplicht met een risicoanalyse.

het ergste geval per agent

Het antwoord op de vraag van je bestuur: wat kan hier misgaan en hoe groot wordt het risico dan?

een to-do-lijst

Alles wat je niet kunt invullen, is een taak.

een gesprek dat ergens over gaat

Met je eigen team, met je leverancier of met je IT-partner.

over ilionx

Wij combineren voor security het beste van het beste, SIEM van Exabeam, netwerkdetectie van Corelight, endpoint- en managed response op SentinelOne, vulnerability management met Tenable en geautomatiseerd pentesten met Horizon3 NodeZero. Ons 24/7 Nederlandse SOC kijkt over die lagen heen mee, met eigen analisten die direct bellen als het kritiek is.

Voor AI-agents werken we samen met Noma. Hun platform vindt de agents in je organisatie, ook de agents die niemand heeft aangemeld en zet rechten op least privilege. Het maakt ook de gevaarlijke combinaties zichtbaar die in losse rechtenlijsten onzichtbaar blijven. Wij verbinden die inzichten met ons eigen SOC, zodat wat Noma ziet ook daadwerkelijk bij iemand terechtkomt die er dezelfde dag iets mee doet.

verantwoording

wetgeving

- Cyberbeveiligingswet en Wet weerbaarheid kritieke entiteiten in werking per 15 augustus 2026, circa 8.000 organisaties, bestuur keurt maatregelen goed en houdt toezicht, Rijksoverheid en NCSC.
- Meldtermijnen 24 uur, 72 uur en één maand, bij CSIRT en toezichthouder, NCSC.
- Deployer-verplichtingen EU AI Act van toepassing per 2 augustus 2026, waaronder menselijk toezicht en logbewaring van minimaal zes maanden, artikel 26 AI Act.

cijfers

- Gemiddelde breakout time 29 minuten in 2025, 65% korter dan 2024, snelst waargenomen 27 seconden, AI-ondersteunde operaties +89%, CrowdStrike Global Threat Report 2026.
- 81% van de teams voorbij de planfase, 14,4% met volledige securitygoedkeuring, 47,1% van agents actief gemonitord, 21,9% met eigen identiteit, 900+ respondenten, Gravitee, State of AI Agent Security 2026 (februari 2026). Gravitee ondervroeg 900+ organisaties over de status van hun AI-agents. Geen trechter, vier losse signalen, niet per se subsets van elkaar.
- 69% deelt ergens inloggegevens tussen agents, 54% had een incident of bijna-incident, 107 organisaties, VentureBeat Pulse Research, juni 2026.

begrippen

- Agent = model + harness; de harness regelt geheugen, tools, rechten en validatie, het model doet dat niet, gangbare definitie in de agent-architectuurpraktijk.
- De vijf eigenschappen en het begrip “toxic combination”, Cloud Security Alliance, Toxic Combinations: The Five Powers Fueling the Agentic Threat Landscape (mei 2026).

Vraag een gesprek aan



Anne van Gemert
Accountmanager
+31 6 50 03 88 76
avangemert@ilionx.com

www.ilionx.com