Normenkader AI-agent assurance

Versie 1.1 (20 november 2026), 49 normen in zes domeinen. Carlo Bavius RE RO CRISC CRMA CIA, Audirium

Dit is de volledige normenlijst bij het artikel «Hoe toets je een AI-agent?». De domeinen volgen de levensloop van een agent: van inventaris en identiteit tot naleving en rapportage. Per norm staan de norm zelf, een toetsaanpak voor de auditor en de bron waarop de norm aansluit, met een risicogewicht (H, M of L).

Het kader sluit aan op de AI-verordening (artikel 4 en 50), ISO/IEC 42001:2023, het IIA AI Auditing Framework en het NIST AI RMF. Ervaringen met toepassing op een andere agent zijn welkom bij de auteur, via carlo@audirium.net.

A. Inventaris en identiteit 8 normen

A01 Compleetheid van de agent-inventaris

De organisatie houdt een actueel en centraal register bij van alle AI-agents die in productie draaien of daarvoor zijn vrijgegeven, ongeacht of ze zelfgebouwd zijn, ingekocht, of als functie ingebed zitten in een SaaS-product. Het register bevat per agent minimaal een unieke naam of code, de systemen waarmee de agent verbinding maakt, en het platform of de omgeving waarop hij draait. Een agent die niet in dit register staat, mag geen productiedata of productiesystemen aanraken.

Toetsaanpak. Vraag het register op en vergelijk het steekproefsgewijs met een onafhankelijke bron, bijvoorbeeld de lijst API-sleutels, service-accounts of geïnstalleerde plug-ins bij drie tot vijf afdelingen. Vraag door op de laatste bijwerkdatum en wie het register onderhoudt.

Afbakening. A01 toetst het register zelf: het bestaat, is actueel, heeft een beheerder en bevat per agent de verplichte velden, en een niet-geregistreerde agent krijgt geen productietoegang; het vinden van niet-gemelde agents hoort bij A02.

Bron: Sluit aan op IIA AI Auditing Framework, domein Governance, aangescherpt voor agents; risicogewicht H; indeling: aanscherping; ISO 27001 bijlage A: 5.9 Inventarisatie van bedrijfsmiddelen

A02 Actieve opsporing van schaduw-agents

De organisatie beperkt zich niet tot vrijwillige melding maar spoort periodiek AI-agents op die buiten het formele traject zijn ingezet, bijvoorbeeld agents die medewerkers zelf hebben gebouwd met low-code platformen, browserextensies met agentfunctionaliteit, of automatiseringen ingebed in ingekochte software die niemand heeft gemeld. De opsporingsmethode is vastgelegd en wordt met een vaste frequentie uitgevoerd. Gevonden schaduw-agents worden binnen een vastgestelde termijn geregistreerd, stopgezet of expliciet geaccepteerd met een risicoafweging.

Toetsaanpak. Vraag de laatste twee opsporingsrondes op inclusief bevindingen en opvolging. Toets zelf één technische bron, bijvoorbeeld netwerkverkeer naar bekende LLM-API-endpoints, tegen het register om te zien of alles daarin al bekend was.

Afbakening. A02 toetst de opsporing: een vastgelegde methode met vaste frequentie, aantoonbare opvolging van gevonden agents en de eigen vergelijking van het register met een onafhankelijke technische bron; inhoud en beheer van het register horen bij A01.

Bron: Eigen norm, agent-specifiek; sluit aan bij het NIST-signaal (februari 2026) dat agents doorgaans ongeregistreerd draaien; risicogewicht H; indeling: aanscherping; ISO 27001 bijlage A: 5.9 Inventarisatie van bedrijfsmiddelen, 8.16 Monitoren van activiteiten

A03 Eigenaarschap per agent

Elke geregistreerde agent heeft een met naam genoemde functionaris of team dat verantwoordelijk is voor het functioneren, de risico's en het buiten gebruik stellen van die agent. Deze eigenaar is niet de leverancier en niet een generiek IT-team, maar iemand binnen de business die het doel van de agent kent en aanspreekbaar is bij incidenten. Wisseling van eigenaar, bijvoorbeeld bij uitdiensttreding, wordt binnen een vaste termijn in het register bijgewerkt.

Toetsaanpak. Neem een steekproef van agents uit het register en verifieer bij de genoemde eigenaar of deze zich bewust is van het eigenaarschap en de bijbehorende verantwoordelijkheid kan uitleggen. Controleer of een recent vertrokken medewerker nog als eigenaar geregistreerd staat.

Afbakening. A03 gaat over de doorlopend aanspreekbare eigenaar in de business en het bijwerken daarvan bij wisseling; het eenmalige besluit waarmee de agent is ingesteld hoort bij B01.

Bron: Sluit aan op IIA AI Auditing Framework, domein Governance; risicogewicht M; indeling: aanscherping; ISO 27001 bijlage A: 5.2 Taken en verantwoordelijkheden voor informatiebeveiliging, 5.9 Inventarisatie van bedrijfsmiddelen

A04 Doel, reikwijdte en herkomst per agent vastgelegd

Van elke agent is schriftelijk vastgelegd waarvoor hij is bedoeld, welke taken en beslissingen binnen zijn reikwijdte vallen, en waar hij vandaan komt: zelf ontwikkeld, ingekocht als losstaand product, of als functie ingebed in een SaaS-toepassing die de organisatie al gebruikte. Bij ingekochte of ingebedde agents is vastgelegd welke leverancier verantwoordelijk is voor het onderliggende model en welke afspraken daarover bestaan. Een agent die taken uitvoert buiten de vastgelegde reikwijdte is een afwijking die gemeld moet worden.

Toetsaanpak. Vraag voor een steekproef agents de doelbeschrijving op en leg die naast een waarneming van de daadwerkelijke acties, bijvoorbeeld uit logging of een demonstratie. Beoordeel of de reikwijdte concreet genoeg is om een overschrijding te kunnen herkennen.

Bron: Eigen norm, agent-specifiek; raakt ISO/IEC 42001 hoofdstuk 8; risicogewicht M; indeling: aanscherping; ISO 27001 bijlage A: 5.9 Inventarisatie van bedrijfsmiddelen

A05 Eigen technische identiteit per agent

Elke agent authenticeert zich met een eigen technische identiteit, bijvoorbeeld een eigen service-account, API-sleutel of certificaat, en draait niet onder een gedeeld of generiek serviceaccount samen met andere agents of processen. Rechten en toegang worden aan die individuele identiteit gekoppeld, niet aan een gedeeld account met bredere rechten dan de agent nodig heeft. Waar een gedeeld account onvermijdelijk is, bijvoorbeeld bij een verouderd platform, is dat expliciet als afwijking vastgelegd met een compenserende maatregel.

Toetsaanpak. Vraag voor een steekproef agents de technische configuratie op (service-account, sleutel of certificaat) en verifieer in het identity- en accessbeheersysteem dat dit een uniek, aan de agent gekoppeld account is en geen gedeeld account. Vraag naar het aantal agents dat nog onder een gedeeld account draait.

Afbakening. A05 is de enige norm over de technische identiteit: een eigen account, sleutel of certificaat per agent en vastgelegde afwijkingen bij gedeelde accounts; wie verantwoordelijk is of wie de agent heeft geautoriseerd hoort bij A03 en B01.

Bron: Eigen norm, agent-specifiek; sluit aan bij de eis van Singapore (IMDA, januari 2026) voor een verifieerbare digitale identiteit per agent; risicogewicht H; indeling: al gedekt; ISO 27001 bijlage A: 5.16 Identiteitsbeheer, 5.18 Toegangsrechten

A06 Koppeling tussen agent-identiteit en logging

Logging van agentacties bevat de unieke technische identiteit van de uitvoerende agent, zodat een handeling altijd herleidbaar is tot een specifieke agent en niet slechts tot een gedeeld account of een generieke applicatienaam. Deze koppeling geldt zowel voor acties die de agent zelf uitvoert als voor acties die hij namens een gebruiker uitvoert, waarbij beide identiteiten apart herkenbaar zijn. Zonder deze koppeling is achteraf niet vast te stellen welke agent verantwoordelijk was voor een handeling.

Toetsaanpak. Vraag een voorbeeld van logregels op voor een geregistreerde agent en controleer of de agent-identiteit als apart, doorzoekbaar veld aanwezig is naast eventuele gebruikersidentiteit. Probeer voor een concrete actie terug te herleiden welke agent deze heeft uitgevoerd.

Afbakening. A06 toetst alleen wie: elke logregel bevat de unieke agent-identiteit en, bij handelen namens een gebruiker, diens identiteit als apart doorzoekbaar veld; wat er verder in de regel moet staan en of de logging volledig is hoort bij D01.

Bron: Eigen norm, agent-specifiek; sluit aan bij de eis van Singapore (IMDA, januari 2026) voor een spoor van welke agent onder wiens machtiging handelde; risicogewicht H; indeling: aanscherping; ISO 27001 bijlage A: 5.16 Identiteitsbeheer, 8.15 Logging

A07 In dienst nemen en buiten gebruik stellen van een agent

Er bestaat een vastgelegde procedure voor het in productie nemen van een nieuwe agent, inclusief registratie, toewijzing van een eigenaar en toekenning van een eigen technische identiteit voordat de agent toegang krijgt tot productiesystemen. Er bestaat eveneens een procedure voor het buiten gebruik stellen van een agent, waarbij toegangsrechten en de technische identiteit worden ingetrokken en de agent uit het register wordt afgevoerd of als inactief gemarkeerd. Een agent die niet langer wordt gebruikt maar nog wel toegang heeft, is een openstaand risico dat via deze procedure zichtbaar moet worden.

Toetsaanpak. Neem een steekproef van recent geïntroduceerde en recent uitgefaseerde agents en volg het dossier tegen de vastgelegde procedure. Controleer voor een uitgefaseerde agent of de bijbehorende technische identiteit daadwerkelijk is ingetrokken, bijvoorbeeld door te verifiëren dat het account niet meer kan authenticeren.

Afbakening. A07 omvat de hele levenscyclus: de procedure voor in productie nemen en voor buiten gebruik stellen, inclusief het daadwerkelijk intrekken van rechten en identiteit wanneer de agent stopt of zijn doel vervalt.

Bron: Eigen norm, agent-specifiek; risicogewicht M; indeling: aanscherping; ISO 27001 bijlage A: 5.9 Inventarisatie van bedrijfsmiddelen, 5.16 Identiteitsbeheer, 5.18 Toegangsrechten

A08 Begrensde en herleidbare delegatieketen

Waar een agent taken doorgeeft aan een andere agent of externe dienst, is vastgelegd welke ontvangers hij mag aanroepen, tot welke diepte, en onder welke machtiging: die van de aanroeper, de oorspronkelijke gebruiker of een eigen beperkte machtiging. Aanroepen buiten deze paden worden geweigerd. Per stap is herleidbaar welke agent namens welk mandaat handelde.

Toetsaanpak. Vraag de vastgelegde delegatiepaden, maximale diepte en machtigingslogica op. Reconstrueer uit de logging een keten van minimaal twee stappen en stel per stap agent-identiteit en mandaat vast. Probeer in test een niet-toegestane aanroep.

Bron: Eigen norm, agent-specifiek; kernonderscheid van dit toetsinstrument ten opzichte van bestaande AI-systeemkaders; opgenomen uit B07: maximale ketendiepte en externe diensten; risicogewicht H; indeling: nieuw; ISO 27001 bijlage A: 5.15 Toegangsbeveiliging, 5.16 Identiteitsbeheer, 5.18 Toegangsrechten, 8.15 Logging

B. Mandaat en begrenzing 9 normen

B01 Vastgelegd mandaat per agent

Voor elke agent is vastgelegd onder wiens autorisatie hij handelt: welke functionaris of proceseigenaar de agent heeft ingesteld, voor welk doel, en op basis van welk besluit. De agent heeft een eigen, herleidbare identiteit en draait niet onder een gedeeld generiek serviceaccount waarachter meerdere agents of gebruikers schuilgaan.

Toetsaanpak. Vraag het register van agents op en toets voor een steekproef of per agent een proceseigenaar, doel en autorisatiebesluit is vastgelegd. Controleer in de infrastructuur (IAM, service-accounts) of de agent een eigen identiteit heeft of dat meerdere agents een gedeeld account delen.

Afbakening. B01 gaat alleen over het mandaat: het vastgelegde besluit met opdrachtgever, doel en datum waarmee de agent is ingesteld; de zin over eigen identiteit en gedeelde accounts vervalt hier en hoort bij A05, het lopende eigenaarschap bij A03.

Bron: NIST (februari 2026): agents draaien doorgaans als generiek serviceaccount zonder eigen identiteit, autorisatie of verantwoording; Singapore IMDA (januari 2026): eis van verifieerbare digitale identiteit per agent; risicogewicht H; indeling: aanscherping; ISO 27001 bijlage A: 5.2 Taken en verantwoordelijkheden voor informatiebeveiliging, 5.16 Identiteitsbeheer, 5.18 Toegangsrechten

B02 Minimale rechten toegepast op agents

De rechten van een agent zijn afgeleid van de taak die hij uitvoert en niet ruimer dan daarvoor nodig is. Rechten die bij het instellen van de agent gemakshalve zijn meegegeven maar niet worden gebruikt, zijn periodiek beoordeeld en ingetrokken.

Toetsaanpak. Vergelijk voor een steekproef van agents de toegewezen rechten (rollen, scopes, sleutels) met de daadwerkelijk uitgevoerde handelingen in de logging over een representatieve periode. Vraag op wanneer de laatste rechtenbeoordeling heeft plaatsgevonden en wat daarvan de uitkomst was.

Afbakening. B02 toetst de omvang van de toegekende rechten tegen het feitelijke gebruik en de periodieke beoordeling en intrekking van ongebruikte rechten; welke systemen, gegevens en gereedschappen binnen de grens vallen hoort bij B03 en B04.

Bron: Aansluitend bij ISO/IEC 42001 (rollen en verantwoordelijkheden in het AI-managementsysteem, niet-letterlijk overgenomen); NIST februari 2026; risicogewicht H; indeling: al gedekt; ISO 27001 bijlage A: 5.18 Toegangsrechten

B03 Begrensde toegang tot systemen en gegevens

Voor elke agent is expliciet vastgelegd welke systemen, databronnen en gegevenscategorieën hij mag benaderen, met bijzondere aandacht voor gevoelige gegevens zoals persoonsgegevens, financiële gegevens of vertrouwelijke bedrijfsinformatie. Toegang buiten deze vastgelegde grens wordt technisch afgedwongen en niet alleen aan de agent opgedragen als instructie.

Toetsaanpak. Vraag de toegangsconfiguratie (allowlist, netwerksegmentatie, databaserechten) op voor een steekproef agents en toets deze tegen de vastgelegde scope. Doe een technische waarneming of het buiten de scope komen daadwerkelijk wordt geweigerd, in plaats van alleen in een prompt of richtlijn te staan.

Afbakening. B03 toetst de vastgelegde grens voor systemen, databronnen en gegevenscategorieën en dat toegang daarbuiten technisch wordt geweigerd; gereedschappen en API's horen bij B04, de vraag of rechten binnen de grens te ruim zijn bij B02.

Bron: NIST februari 2026 (ontbrekende autorisatie bij generieke serviceaccounts); risicogewicht H; indeling: aanscherping; ISO 27001 bijlage A: 5.15 Toegangsbeveiliging, 8.3 Beperking van toegang tot informatie, 8.22 Segregatie van netwerken

B04 Toegestane en verboden gereedschappen vastgelegd

Per agent is een lijst vastgelegd van gereedschappen, API's en plug-ins die hij mag aanroepen, met expliciete benoeming van categorieën die verboden zijn, zoals ongefilterde internettoegang, code-executie op productiesystemen of ongedocumenteerde API's. Een nieuw gereedschap wordt pas beschikbaar gesteld na een expliciete beoordeling.

Toetsaanpak. Vraag het toolregister van een steekproef agents op en vergelijk dit met de daadwerkelijk beschikbare functies in de technische implementatie, bijvoorbeeld de tool-definities in de agent-omgeving. Controleer of er een wijzigingsproces bestaat voor het toevoegen van een gereedschap en of dat proces is gevolgd bij een recente uitbreiding.

Afbakening. B04 toetst de lijst van toegestane en verboden gereedschappen, API's en plug-ins en de beoordeling vóór toevoeging van een nieuw gereedschap; tot welke gegevens een gereedschap toegang geeft hoort bij B03.

Bron: Eigen normstelling, agent-specifiek; risicogewicht M; indeling: aanscherping; ISO 27001 bijlage A: 8.18 Gebruik van bevoorrechte hulpprogramma's, 8.32 Wijzigingsbeheer

B05 Menselijke tussenkomst bij onomkeerbare handelingen

Handelingen met een onomkeerbaar of extern gericht karakter, zoals betalen, gegevens verwijderen, extern communiceren namens de organisatie of een verplichting of contract aangaan, worden nooit volledig zelfstandig door een agent uitgevoerd. Voor deze categorieën is een verplicht goedkeuringsmoment door een mens ingebouwd voordat de handeling wordt uitgevoerd.

Toetsaanpak. Vraag de lijst van handelingscategorieën op die als onomkeerbaar of extern gericht zijn aangemerkt en toets of daarvoor een technische goedkeuringsstap is ingebouwd, niet alleen een richtlijn. Test in een niet-productieomgeving of de agent een dergelijke handeling daadwerkelijk kan voltooien zonder die goedkeuring.

Afbakening. B05 toetst uitsluitend de werking: voor de in C01 aangewezen categorieën bestaat een technische goedkeuringsstap die de agent niet kan omzeilen; welke categorieën dat zijn en waarom hoort bij C01.

Bron: Eigen normstelling, aansluitend bij governance-uitgangspunten van het IIA AI Auditing Framework (niet-letterlijk overgenomen); risicogewicht H; indeling: aanscherping; ISO 27001 bijlage A: geen tegenhanger

B06 Limieten op handelingen, uitgaven en doorlooptijd

Voor elke agent gelden kwantitatieve grenzen aan zijn handelen: een maximaal aantal acties per tijdseenheid, een uitgavenplafond waar van toepassing, en een maximale doorlooptijd per taak of sessie. Bij het bereiken van een limiet stopt de agent automatisch of schakelt hij door naar een mens, in plaats van dat de limiet alleen als richtlijn is gedocumenteerd.

Toetsaanpak. Vraag de technische instelling van de limieten op (bijvoorbeeld rate limits, budgetplafonds, timeouts) en toets of deze daadwerkelijk zijn geconfigureerd in de omgeving waarin de agent draait. Vraag logging op van een moment waarop een limiet is bereikt en beoordeel of het vastgelegde vervolggedrag daadwerkelijk is opgetreden.

Bron: Eigen normstelling, agent-specifiek; risicogewicht M; indeling: nieuw; ISO 27001 bijlage A: 8.6 Capaciteitsbeheer

B08 Scheiding test/productie en intrekking bij vervallen doel

Een agent die in een testomgeving is ontwikkeld of beproefd, beschikt niet automatisch over dezelfde rechten of dezelfde toegang tot productiegegevens en productiesystemen als in productie. Wanneer het doel waarvoor een agent is ingesteld vervalt of de agent buiten gebruik wordt gesteld, worden zijn rechten en toegangen actief ingetrokken in plaats van passief te laten voortbestaan.

Toetsaanpak. Vraag op hoe test- en productieomgevingen van agents zijn gescheiden (aparte accounts, aparte data, aparte rechten) en toets dit voor een steekproef. Vraag het register van uitgefaseerde of niet langer gebruikte agents op en controleer, bijvoorbeeld door de laatste gebruiksdatum te vergelijken met de status van hun toegang, of hun rechten daadwerkelijk zijn ingetrokken.

Afbakening. B08 gaat alleen over de scheiding tussen test en productie: een agent in test heeft eigen accounts, eigen data en geen productierechten; het intrekken van rechten bij vervallen doel of uitfasering vervalt hier en hoort bij A07.

Bron: Eigen normstelling, aansluitend bij wijzigingsbeheer in ISO/IEC 42001 (niet-letterlijk overgenomen); risicogewicht M; indeling: al gedekt; ISO 27001 bijlage A: 5.18 Toegangsrechten, 8.31 Scheiding van ontwikkel-, test- en productieomgevingen, 8.33 Testinformatie

B09 Doelbinding van de gegevens die de agent verwerkt

Van elke agent is vastgelegd voor welk doel hij persoonsgegevens verwerkt en is beoordeeld of dat doel verenigbaar is met het doel waarvoor die gegevens oorspronkelijk zijn verzameld. Een agent die een bestaande verzameling voor een nieuwe taak ontsluit, bijvoorbeeld een auditdossier doorzoeken op signalen, vormt een nieuw verwerkingsdoel dat expliciet is beoordeeld en niet stilzwijgend meelift met de toegangsrechten van de gebruiker namens wie de agent handelt. Waar de verwerking een hoog risico voor betrokkenen kan opleveren, is vastgesteld of een gegevensbeschermingseffectbeoordeling nodig is en is die uitgevoerd.

Toetsaanpak. De auditor vraagt per agent het vastgelegde verwerkingsdoel op en vergelijkt dat met het doel van de bronsystemen waaruit de agent put. Voor een steekproef agents wordt nagegaan of de doelbeoordeling dateert van vóór ingebruikname, wie haar heeft vastgesteld, en of bij een hoog-risicoverwerking een effectbeoordeling aanwezig is. Een agent waarvan het doel alleen impliciet blijkt uit de rechten van de aanroepende gebruiker geldt als afwijking.

Bron: AVG (EU) 2016/679 art. 5 lid 1 sub b en art. 35, agent-specifiek toegepast (geen letterlijke overname); risicogewicht H; indeling: vertaling van wet; ISO 27001 bijlage A: 5.34 Privacy en bescherming van persoonsgegevens

B10 Dataminimalisatie en meeliftende persoonsgegevens

De agent krijgt niet meer gegevens te zien dan zijn taak vereist. Waar de taak geen persoonsgegevens nodig heeft, worden die vóór verwerking verwijderd, gepseudonimiseerd of gemaskeerd. De organisatie heeft onderkend dat in ongestructureerde bronnen zoals auditdossiers, mailboxen, exports en bijlagen persoonsgegevens meeliften die niemand bewust aan de agent heeft aangeboden, waaronder bijzondere of strafrechtelijke persoonsgegevens, en heeft vastgesteld hoe zij die herkent en behandelt. Een inrichting waarin de agent eenvoudigweg bij alles kan wat de aanroepende gebruiker kan, geldt niet als minimalisatie.

Toetsaanpak. De auditor neemt een steekproef van daadwerkelijk aan de agent aangeboden invoer, bijvoorbeeld uit de promptlogging, en stelt vast welke persoonsgegevens daarin voorkomen en of die voor de taak nodig waren. Getoetst wordt of er een maatregel vooraf is (filtering, maskering, pseudonimisering) en of die werkt, en of de organisatie bijzondere of strafrechtelijke gegevens in ongestructureerde bronnen actief herkent in plaats van erop te vertrouwen dat ze er niet in zitten.

Bron: AVG (EU) 2016/679 art. 5 lid 1 sub c, art. 9 en art. 10; aansluitend bij ISO/IEC 27001:2022 bijlage A 8.11 (niet-letterlijk overgenomen); risicogewicht H; indeling: vertaling van wet; ISO 27001 bijlage A: 5.34 Privacy en bescherming van persoonsgegevens, 8.3 Beperking van toegang tot informatie, 8.11 Datamaskering

C. Menselijke tussenkomst 7 normen

C01 Goedkeuringsdrempel voor impactvolle handelingen

De organisatie heeft vastgelegd voor welke categorieën handelingen van de agent voorafgaande menselijke goedkeuring verplicht is, bijvoorbeeld betalingen boven een bedrag, wijzigingen aan productiedata, communicatie naar buiten of het verwijderen van gegevens. De drempel is onderbouwd op basis van omkeerbaarheid en impact van de handeling, niet op basis van wat technisch het makkelijkst uit te bouwen was, en is actueel ten opzichte van de daadwerkelijke bevoegdheden van de agent.

Toetsaanpak. Auditor vraagt de lijst van goedkeuringsplichtige handelingen op, legt die naast de daadwerkelijke tool- en actiebevoegdheden van de agent, en toetst een steekproef van recent uitgevoerde handelingen op of ze onder de juiste drempel vielen.

Afbakening. C01 toetst uitsluitend de opzet: de lijst goedkeuringsplichtige handelingen bestaat, is onderbouwd op omkeerbaarheid en impact en dekt de actuele bevoegdheden van de agent; of de goedkeuringsstap technisch afdwingt hoort bij B05.

Bron: Sluit aan bij de NIST-bevinding (februari 2026) dat agents doorgaans zonder eigen autorisatiekader draaien; eigen formulering vanuit agent-perspectief.; risicogewicht H; indeling: aanscherping; ISO 27001 bijlage A: geen tegenhanger

C02 Aantoonbaarheid van een echte beoordeling

De organisatie kan aantonen dat een goedkeuring een inhoudelijke beoordeling was en geen automatisch doorklikken. Dit blijkt bijvoorbeeld uit gemeten beoordelingstijd, uit een steekproef van goedkeuringen op onderbouwing, of uit monitoring van het percentage goedkeuringen zonder enige aanpassing of afwijzing als signaal voor waakzaamheidsverlies. Een structureel goedkeuringspercentage richting de honderd procent wordt als risico erkend en opgevolgd, niet stilzwijgend geaccepteerd als bewijs dat het goed gaat.

Toetsaanpak. Auditor vraagt logging van goedkeuringstijden en -percentages op over een representatieve periode, beoordeelt of er een signaleringsmechanisme voor massale of te snelle goedkeuring bestaat, en bevraagt de proceseigenaar op de opvolging wanneer dat signaal afgaat.

Bron: Eigen norm, gericht op automation bias bij agentgoedkeuring; dit specifieke risico is in het IIA-framework en ISO 42001 niet op agent-niveau uitgewerkt.; risicogewicht H; indeling: nieuw; ISO 27001 bijlage A: geen tegenhanger

C03 Voldoende informatie voor de beoordelaar

De mens die een voorgestelde handeling van de agent moet goedkeuren, krijgt daarbij de onderliggende overweging, de gebruikte brondata en het verwachte effect te zien, niet alleen een samenvattend verzoek om akkoord. De hoeveelheid en vorm van die informatie is afgestemd op wat nodig is om binnen de beschikbare tijd een verantwoorde beslissing te nemen.

Toetsaanpak. Auditor laat zich een aantal goedkeuringsverzoeken tonen of reconstrueert deze uit logs, en beoordeelt of een beoordelaar zonder aanvullend eigen onderzoek een verantwoorde ja- of nee-beslissing had kunnen nemen.

Bron: Eigen norm vanuit de agent-invalshoek; raakt aan artikel 4 AI-verordening (AI-geletterdheid) voor zover het de bekwaamheid van de beoordelaar betreft om de getoonde informatie te begrijpen.; risicogewicht M; indeling: aanscherping; ISO 27001 bijlage A: geen tegenhanger

C04 Bevoegdheid en snelheid van de noodstop

Vastgelegd is wie de agent mag stopzetten, benoemd als rol en niet als individuele naam, via welk kanaal dit gebeurt, en binnen welke tijd de stop daadwerkelijk doorwerkt op alle lopende en nieuwe handelingen van de agent. De stopbevoegdheid is niet afhankelijk van de beschikbaarheid van de bouwer of leverancier van de agent.

Toetsaanpak. Auditor vraagt de procedure en het technisch mechanisme van de noodstop op, controleert of de bevoegde rol daadwerkelijk toegang heeft tot dat mechanisme, en toetst de vastgelegde reactietijd tegen wat technisch aantoonbaar haalbaar is.

Afbakening. C04 toetst de regeling: welke rol mag stoppen, via welk kanaal, met welke toegezegde doorwerktijd en zonder afhankelijkheid van bouwer of leverancier; of die tijd in de praktijk wordt gehaald hoort bij C05.

Bron: Sluit aan bij de Singapore IMDA-eis (januari 2026) voor traceerbare machtiging per agent; eigen formulering.; risicogewicht H; indeling: aanscherping; ISO 27001 bijlage A: 5.24 Planning en voorbereiding van incidentbeheer, 5.26 Respons op informatiebeveiligingsincidenten

C05 Beproeving van de noodstop en afhandeling van onderhanden werk

De noodstop is minstens een keer daadwerkelijk beproefd en niet alleen op papier beschreven, en het resultaat van die beproeving is vastgelegd. Er is bepaald wat er met een handeling gebeurt die de agent halverwege uitvoert op het moment van de stop: veilig afmaken, annuleren, of achterlaten in een herstelbare tussentoestand.

Toetsaanpak. Auditor vraagt bewijs op van de laatst uitgevoerde noodstoptest, inclusief datum, resultaat en eventuele bijstellingen naar aanleiding daarvan, en laat zich aan de hand van een concreet voorbeeld uit logs toelichten hoe halfafgemaakt werk is afgehandeld.

Afbakening. C05 toetst de werking: de noodstop is daadwerkelijk beproefd, de gemeten doorwerktijd is vastgelegd naast de toezegging uit C04, en bepaald is wat er met een halverwege onderbroken handeling gebeurt.

Bron: Eigen norm, gericht op de agent-specifieke risico's van tussentoestanden in een keten van achtereenvolgende tool-aanroepen; niet gedekt door bestaande AI-systeemkaders.; risicogewicht H; indeling: aanscherping; ISO 27001 bijlage A: geen tegenhanger

C06 Terugdraaien van reeds uitgevoerde handelingen

Voor handelingen die de agent zelfstandig mag uitvoeren, is bepaald of en hoe deze achteraf kunnen worden teruggedraaid of gecompenseerd, bijvoorbeeld een verzonden bericht, een geplaatste order of een gewijzigd record, inclusief wie hierover beslist en binnen welke termijn herstel nog mogelijk is. Waar terugdraaien niet mogelijk is, is dat vooraf onderkend en meegewogen in de goedkeuringsdrempel uit norm C01.

Toetsaanpak. Auditor kiest een aantal moeilijk omkeerbare handelingstypen uit de bevoegdhedenlijst, vraagt per type na hoe herstel is geregeld en of dit ooit is toegepast, en vergelijkt de uitkomst met wat vooraf was voorzien.

Afbakening. C06 gaat alleen over handelingen die de agent zonder voorafgaande goedkeuring mag uitvoeren: hoe, door wie en binnen welke termijn ze worden teruggedraaid of gecompenseerd; het goedkeuringsmoment vooraf hoort bij C01 en B05.

Bron: Eigen norm vanuit de agent-invalshoek, aansluitend bij het onderscheid omkeerbaar/onomkeerbaar dat bij de bevoegdheden van de agent wordt vastgesteld.; risicogewicht M; indeling: nieuw; ISO 27001 bijlage A: geen tegenhanger

C07 Escalatie bij twijfel en bereikbaarheid van de verantwoordelijke

De agent escaleert een geval buiten zijn vastgestelde bereik of met een lage zekerheidsmarge naar een mens, in plaats van zelf een keuze te forceren. De organisatie heeft geregeld dat tijdens de daadwerkelijke draaiuren van de agent een bevoegde mens bereikbaar is om zo'n escalatie op te pakken, en heeft vastgelegd wat er buiten die uren gebeurt: de agent staat stil, wacht, of werkt door onder een ruimere vooraf bepaalde marge.

Toetsaanpak. Auditor vraagt de escalatiecriteria en -route op, toetst de bezettings- of bereikbaarheidsregeling tegen de daadwerkelijke draaitijden van de agent, en beoordeelt aan de hand van een eerdere escalatie of deze binnen een redelijke termijn is opgepakt.

Bron: Eigen norm, gericht op het gat dat NIST signaleert tussen continu draaiende agents en generieke, niet-geïndividualiseerde menselijke supervisie.; risicogewicht M; indeling: aanscherping; ISO 27001 bijlage A: geen tegenhanger

D. Bewijs en herleidbaarheid 9 normen

D01 Logging van elke agenthandeling

De organisatie legt van elke handeling van een agent vast: wat de handeling was, op welk tijdstip, onder welk mandaat (namens welke gebruiker of welk proces) en met welk resultaat of gevolg. Deze vastlegging gebeurt automatisch en is niet uitzetbaar door de agent zelf of door degene die de agent bedient. Voor agents die transacties uitvoeren of gegevens wijzigen is de vastlegging volledig, niet steekproefsgewijs.

Toetsaanpak. De auditor selecteert een reeks agent-acties uit een productieperiode en vraagt de bijbehorende logregels op; hij toetst of wat, wanneer, mandaat en gevolg per actie aanwezig zijn en of er gaten in de reeks zitten.

Afbakening. D01 eist dat elke handeling automatisch, volledig en niet-uitzetbaar wordt vastgelegd met wat, wanneer en resultaat; welk identiteitsveld de regel draagt hoort bij A06, de aansluiting op het bronsysteem bij D08 en het reconstrueren voor derden bij F07.

Bron: NIST (februari 2026): agents draaien doorgaans als generieke serviceaccounts zonder eigen identiteit, autorisatie of verantwoording; Singapore IMDA (januari 2026) eist een spoor van welke agent onder wiens machtiging handelde.; risicogewicht H; indeling: aanscherping; ISO 27001 bijlage A: 8.15 Logging

D02 Herleidbaarheid naar instructie en modelversie

Van elke uitspraak, aanbeveling of elk besluit van een agent is vast te stellen welke instructie (systeemprompt of taakopdracht) en welke modelversie deze heeft voortgebracht. Zonder die koppeling kan de organisatie een resultaat niet reproduceren of verklaren, en kan zij niet vaststellen of een fout aan het model, de instructie of de invoer lag.

Toetsaanpak. De auditor kiest een agent-output uit het verleden en vraagt de organisatie deze te herleiden naar de exacte instructie en modelversie; hij beoordeelt of dit binnen redelijke tijd en zonder giswerk lukt.

Afbakening. D02 toetst de koppeling per uitkomst: bij elke output is vastgelegd welke instructieversie en welke modelversie hem voortbracht; of die instructieversies zelf bewaard en toegeschreven zijn hoort bij D03.

Bron: Aansluitend op de traceerbaarheidsgedachte van het IIA AI Auditing Framework, agent-specifiek aangescherpt met de koppeling naar instructie en modelversie.; risicogewicht H; indeling: nieuw; ISO 27001 bijlage A: 8.9 Configuratiebeheer, 8.15 Logging

D03 Versiebeheer van agentinstructies

Instructies en prompts waarmee een agent wordt aangestuurd staan onder versiebeheer: elke wijziging is gedateerd, herleidbaar naar de wijzigende persoon of het wijzigende proces, en oudere versies blijven raadpleegbaar. Hierdoor is een oordeel van een agent van drie maanden geleden nog te reconstrueren met de instructie zoals die toen luidde, niet zoals die nu luidt.

Toetsaanpak. De auditor vraagt de instructieversie op die drie maanden geleden gold voor een specifieke agent, vergelijkt deze met de huidige versie en controleert of het verschil zichtbaar, gedateerd en toegeschreven is.

Afbakening. D03 toetst alleen het versiebeheer van instructies en prompts: elke wijziging gedateerd, toegeschreven en oudere versies raadpleegbaar; de koppeling van een concrete output aan een versie hoort bij D02, het beproeven van een wijziging bij E04.

Bron: Eigen norm, vult het wijzigingsbeheer uit ISO/IEC 42001 aan met agent-specifieke reconstrueerbaarheid van instructies.; risicogewicht M; indeling: aanscherping; ISO 27001 bijlage A: 8.9 Configuratiebeheer, 8.32 Wijzigingsbeheer

D04 Registratie van mislukte agenthandelingen

Mislukte, afgebroken of geweigerde handelingen van een agent worden even consequent vastgelegd als geslaagde handelingen, inclusief de reden van mislukking. Zonder deze registratie is het foutpercentage van een agent onbekend en kan de organisatie geen onderbouwde uitspraak doen over de betrouwbaarheid ervan.

Toetsaanpak. De auditor vraagt de foutlog van een agent op over een vaste periode en toetst deze tegen een onafhankelijke bron, bijvoorbeeld systeemfoutmeldingen of gebruikersklachten, om te zien of mislukkingen daadwerkelijk worden vastgelegd en niet stilzwijgend verdwijnen.

Bron: Eigen norm, gericht op de blinde vlek die NIST (februari 2026) signaleert bij generieke serviceaccounts zonder verantwoording.; risicogewicht H; indeling: aanscherping; ISO 27001 bijlage A: 8.15 Logging, 8.16 Monitoren van activiteiten

D05 Onveranderlijkheid van het agent-logboek

De agent heeft geen schrijf- of wijzigingsrechten op zijn eigen logboek; logregels worden weggeschreven naar een voorziening die de agent niet kan aanpassen of wissen. Personen met beheerrechten op het logboek zijn niet dezelfde personen die de agent operationeel aansturen of configureren.

Toetsaanpak. De auditor vraagt de technische opzet van de logvoorziening op en toetst of het service-account van de agent schrijfrechten heeft op historische logregels; waar mogelijk verifieert hij dit met een technische steekproef.

Bron: Aansluitend op de eis van Singapore IMDA (januari 2026) van een verifieerbaar spoor per agent.; risicogewicht H; indeling: al gedekt; ISO 27001 bijlage A: 5.3 Scheiding van taken, 5.33 Beschermen van registraties, 8.15 Logging

D06 Bewaartermijn afgestemd op verjaring

De bewaartermijn van agent-logs is afgeleid van de verjaringstermijn van het besluit of de handeling die de agent uitvoerde, niet van een generieke IT-standaardtermijn. Voor besluiten met een langere juridische of fiscale verjaring bewaart de organisatie het bijbehorende agent-spoor minstens even lang.

Toetsaanpak. De auditor vergelijkt de vastgestelde bewaartermijn van agent-logs met de verjaringstermijn van het zwaarste besluitstype dat de agent neemt, en controleert of oudere, nog relevante logs niet al zijn verwijderd.

Bron: Eigen norm.; risicogewicht M; indeling: aanscherping; ISO 27001 bijlage A: 5.33 Beschermen van registraties

D07 Eerlijkheid van getoonde zekerheidsindicaties

Een zekerheidspercentage, score of andere indicatie die een agent aan de gebruiker toont, is aantoonbaar gerelateerd aan de daadwerkelijke juistheid van de uitkomst. De organisatie kan uitleggen hoe de indicatie tot stand komt en heeft dit getoetst, zodat de indicatie geen schijnzekerheid geeft die losstaat van de werkelijke betrouwbaarheid.

Toetsaanpak. De auditor vraagt uitleg over de herkomst van een getoonde zekerheidsindicatie en toetst een steekproef van uitkomsten met een hoge en een lage indicatie tegen de daadwerkelijke juistheid ervan.

Afbakening. D07 geldt alleen waar de agent een zekerheidsindicatie aan de gebruiker toont en toetst of die indicatie aantoonbaar samenhangt met de werkelijke juistheid; hoe vaak de agent fout zit hoort bij E02 en E07.

Bron: Eigen norm, aansluitend bij de transparantieverplichting van artikel 50 AI-verordening, hier doorgetrokken naar de betrouwbaarheid van wat aan de gebruiker getoond wordt.; risicogewicht M; indeling: nieuw; ISO 27001 bijlage A: geen tegenhanger

D08 Aansluiting agent-logboek op bronsysteem

Elke handeling die het agent-logboek registreert, is terug te vinden in het bronsysteem waarin de handeling daadwerkelijk landde, bijvoorbeeld een boeking, een ticket of een wijziging in een dossier. De organisatie heeft vastgesteld dat er geen handelingen in het bronsysteem plaatsvinden die niet in het agent-logboek terugkomen, en omgekeerd.

Toetsaanpak. De auditor selecteert onafhankelijk van elkaar een steekproef handelingen uit het bronsysteem en uit het agent-logboek, sluit ze een op een op elkaar aan en onderzoekt afwijkingen op oorzaak.

Afbakening. D08 toetst alleen de tweezijdige aansluiting: elke gelogde handeling is terug te vinden in het bronsysteem en elke agenthandeling in het bronsysteem in het logboek, met verklaarde verschillen; welke velden een logregel bevat hoort bij D01 en A06.

Bron: Eigen norm, sluit aan bij de traditionele auditpraktijk van brontransactie-aansluiting, hier agent-specifiek toegepast.; risicogewicht H; indeling: nieuw; ISO 27001 bijlage A: 8.15 Logging

D09 Persoonsgegevens in prompts, geheugen en afgeleide opslag

De organisatie weet welke persoonsgegevens buiten het bronsysteem achterblijven: in prompts en promptlogging, in het korte- en langetermijngeheugen van de agent, in vector- of zoekindexen die uit brondocumenten zijn opgebouwd, en in tussenopslag bij de leverancier. Voor elk van die plaatsen is een bewaartermijn vastgesteld en bestaat een verwijderroute die daadwerkelijk is beproefd, zodat een verwijdering in het bronsysteem niet stilzwijgend ongedaan wordt gemaakt door een kopie in een index of in het geheugen van de agent.

Toetsaanpak. De auditor laat zich de volledige lijst opslagplaatsen tonen waar invoer van de agent terechtkomt en vergelijkt die met de technische werkelijkheid, bijvoorbeeld door voor één document na te gaan waar het overal is beland. Vervolgens wordt één verwijderverzoek end-to-end gevolgd: is het document ook uit de vectorindex, het agentgeheugen en de promptlogging verdwenen, en binnen welke termijn. Een beschreven maar nooit uitgevoerde verwijderroute is onvoldoende.

Afbakening. D09 gaat over kopieën binnen de eigen agentomgeving: promptlogging, agentgeheugen en vector- of zoekindexen, elk met bewaartermijn en beproefde verwijderroute; opslag en verzoeken bij leverancier, subverwerkers en modelaanbieders horen bij F09.

Bron: AVG (EU) 2016/679 art. 5 lid 1 sub e en art. 17; aansluitend bij ISO/IEC 27001:2022 bijlage A 8.10 (niet-letterlijk overgenomen); risicogewicht H; indeling: vertaling van wet; ISO 27001 bijlage A: 5.34 Privacy en bescherming van persoonsgegevens, 8.10 Informatie verwijderen

E. Betrouwbaarheid en toetsing 7 normen

E01 Beproevingsnorm voorafgaand aan ingebruikname

Voordat een agent taken voor echte gebruikers of op productiedata mag uitvoeren, is er een vooraf vastgestelde norm waartegen zijn gedrag is beproefd: welke taken hij moet aankunnen, welke uitkomst daarbij goed is en welke score minimaal nodig is om vrij te geven. De vrijgave is een expliciet besluit van een verantwoordelijke, niet het stilzwijgend live gaan van een doorontwikkelde versie.

Toetsaanpak. Vraag het beproevingsrapport en het vrijgavebesluit op van de meest recente ingebruikname en vergelijk de daadwerkelijk behaalde score met de vooraf vastgestelde norm en met wie akkoord heeft gegeven.

Bron: Eigen normering, raakt aan ISO 42001 (prestatie-evaluatie) en IIA AI Auditing Framework (validatie voorafgaand aan inzet); risicogewicht H; indeling: aanscherping; ISO 27001 bijlage A: 8.29 Beveiligingstests bij ontwikkeling en acceptatie, 8.32 Wijzigingsbeheer

E02 Bekend foutpercentage en meetmethode

De organisatie kan voor elke agent een foutpercentage noemen: hoe vaak hij een taak niet, deels of onjuist uitvoert, gemeten op een representatieve set taken met een vastgelegde definitie van wat als fout telt. Zonder die definitie en meting is elke uitspraak over betrouwbaarheid een schatting in plaats van een vaststelling.

Toetsaanpak. Vraag de laatst gemeten foutpercentages op per agent, samen met de onderliggende toetsset en de definitie van 'fout', en beoordeel of de toetsset de daadwerkelijke productietaken dekt.

Afbakening. E02 toetst dat per agent een foutpercentage bestaat op een representatieve toetsset met vastgelegde foutdefinitie; het herhalen van die meting hoort bij E03, het beoordelen van echte productie-uitkomsten bij E07.

Bron: Eigen normering, aansluitend op AI-verordening artikel 15 (nauwkeurigheid) voor zover van toepassing; risicogewicht H; indeling: aanscherping; ISO 27001 bijlage A: geen tegenhanger

E03 Herhaalde meting na ingebruikname

De prestaties van een agent worden na ingebruikname op vaste momenten opnieuw gemeten tegen dezelfde toetsset, ook als er niets aan de agent zelf is gewijzigd. Dit vangt stille achteruitgang op wanneer de agent draait op een onderliggend model dat door de leverancier buiten medeweten van de organisatie is bijgewerkt.

Toetsaanpak. Vraag de meetfrequentie en de laatste drie hermetingen op en toets of de resultaten zijn vergeleken met de eerdere baseline en of een afwijking tot een actie heeft geleid.

Afbakening. E03 toetst alleen de herhaling: dezelfde toetsset op vaste momenten, vergeleken met de baseline en met actie bij afwijking; de opzet van toetsset en foutdefinitie hoort bij E02, hermeting naar aanleiding van een wijziging bij E04.

Bron: Eigen normering, specifiek voor agent-risico (modeldrift buiten wijzigingsbeheer van de organisatie), niet gedekt door ISO 42001 of het IIA-framework; risicogewicht H; indeling: aanscherping; ISO 27001 bijlage A: geen tegenhanger

E04 Regressiebeproeving bij wijziging

Elke wijziging aan het onderliggend model, de systeeminstructie of een gereedschap dat de agent gebruikt, wordt vóór doorvoering beproefd op de bestaande toetsset, zodat een verbetering op één punt geen achteruitgang op een ander punt verbergt. De uitkomst van die regressiebeproeving wordt vastgelegd en gekoppeld aan de wijziging.

Toetsaanpak. Selecteer een recente wijziging aan model, instructie of gereedschap van een agent en vraag het bijbehorende regressierapport op; controleer of het dateert van vóór de doorvoering.

Afbakening. E04 toetst dat een wijziging aan model, instructie of gereedschap vóór doorvoering op de toetsset is beproefd en dat het resultaat aan de wijziging is gekoppeld; het vastleggen en bewaren van de versies zelf hoort bij D03.

Bron: Eigen normering, raakt aan ISO 42001 (wijzigingsbeheer); risicogewicht M; indeling: aanscherping; ISO 27001 bijlage A: 8.29 Beveiligingstests bij ontwikkeling en acceptatie, 8.32 Wijzigingsbeheer

E05 Weerbaarheid tegen vijandige en onverwachte invoer

Er is vastgesteld en beproefd hoe de agent reageert op invoer die instructies bevat die niet van de bevoegde gebruiker komen, zoals tekst uit een binnengekomen document, e-mail of webpagina die de agent probeert om te sturen. De agent is zo ingericht dat hij dergelijke ingebedde instructies herkent of in elk geval niet zonder meer opvolgt.

Toetsaanpak. Vraag het testresultaat van een prompt-injectiebeproeving op (invoer met een verborgen instructie) en beoordeel of de agent die instructie heeft genegeerd of gemeld in plaats van uitgevoerd.

Bron: Eigen normering, aansluitend op NIST-signalering over serviceaccounts zonder eigen identiteit als extra kwetsbaarheid bij dit risico; risicogewicht H; indeling: aanscherping; ISO 27001 bijlage A: 8.26 Beveiligingseisen voor toepassingen, 8.29 Beveiligingstests bij ontwikkeling en acceptatie

E06 Scheefheid in uitkomsten die mensen raken

Waar een agent uitkomsten produceert die personen direct raken, zoals een selectie, score, prioritering of afwijzing, is onderzocht of die uitkomsten stelselmatig verschillen tussen groepen mensen zonder inhoudelijke rechtvaardiging. Waar scheefheid is aangetroffen, is vastgelegd welke maatregel is genomen of waarom het risico is geaccepteerd.

Toetsaanpak. Vraag op of er een scheefheidsanalyse is uitgevoerd voor agents met mensgerichte uitkomsten en beoordeel de bevindingen en het vervolg daarop tegen de kenmerken van de betrokken doelgroep.

Bron: Eigen normering, raakt aan AI-verordening (non-discriminatie) en IIA AI Auditing Framework (bias); risicogewicht M; indeling: al gedekt; ISO 27001 bijlage A: geen tegenhanger

E07 Zelfverzekerd onjuist gedrag en herbeoordeling

De organisatie meet niet alleen hoe vaak een agent een taak niet afmaakt, maar ook hoe vaak hij een taak met schijnbare zekerheid afrondt terwijl de uitkomst inhoudelijk onjuist is, omdat dat laatste voor een gebruiker moeilijker te herkennen is. Een deskundige beoordelaar herbeoordeelt periodiek een steekproef van afgeronde agent-uitkomsten, los van de eigen rapportage van de agent over zijn taak.

Toetsaanpak. Vraag de meest recente steekproefsgewijze herbeoordeling op, inclusief hoe vastlopen en zelfverzekerd-onjuist afzonderlijk zijn geteld, en controleer of de beoordelaar onafhankelijk was van de bouw of inzet van de agent.

Afbakening. E07 toetst de periodieke herbeoordeling door een onafhankelijke deskundige van afgeronde productie-uitkomsten, waarbij inhoudelijk onjuist maar afgerond apart van vastlopen wordt geteld; de meting op de toetsset hoort bij E02, een aan de gebruiker getoonde score bij D07.

Bron: Eigen normering, specifiek voor agent-risico (autonoom afgeronde taken zonder vangnet), niet gedekt door ISO 42001 of het IIA-framework; risicogewicht H; indeling: nieuw; ISO 27001 bijlage A: geen tegenhanger

F. Naleving en rapportage 9 normen

F01 Kenbaarheid van AI-agent bij klantcontact

De organisatie heeft per kanaal (chat, e-mail, telefoon) vastgelegd hoe de tegenpartij te weten komt dat zij met een AI-agent communiceert en niet met een mens. Deze kenbaarmaking is zichtbaar of hoorbaar bij de eerste interactie, niet weggestopt in algemene voorwaarden. Voor telefonie is expliciet geregeld hoe wordt gemeld dat een agent het gesprek (mede) voert.

Toetsaanpak. De auditor doorloopt zelf een interactie via elk kanaal waarop een agent actief is en stelt vast of de melding daadwerkelijk verschijnt vóór of bij aanvang van het contact, en vraagt de proceseigenaar om onderbouwing waarom dit toereikend is.

Bron: AI-verordening (EU) 2024/1689, artikel 50 (transparantieplicht, handhaafbaar sinds 2 augustus 2026); risicogewicht H; indeling: vertaling van wet; ISO 27001 bijlage A: geen tegenhanger

F02 Markering van door de agent voortgebrachte content

Content die de agent zelfstandig produceert en extern verspreidt of publiceert, zoals teksten, rapportages, samenvattingen of gegenereerde beelden, is herkenbaar gemarkeerd als machinaal voortgebracht. Dit geldt ook wanneer een medewerker die content ongewijzigd overneemt. De markering gaat niet verloren zodra de content buiten het oorspronkelijke systeem wordt gedeeld of gekopieerd.

Toetsaanpak. De auditor vraagt content op die de agent extern heeft verspreid, controleert of de markering aanwezig en leesbaar is, en test of deze behouden blijft bij kopiëren of doorsturen buiten het systeem.

Bron: AI-verordening (EU) 2024/1689, artikel 50 lid 2 (markering van synthetische content); risicogewicht M; indeling: vertaling van wet; ISO 27001 bijlage A: geen tegenhanger

F03 AI-geletterdheid van gebruikers en beoordelaars

Medewerkers die de agent aansturen of instrueren, en medewerkers die de uitkomsten van de agent beoordelen of goedkeuren, hebben aantoonbaar kennis van de mogelijkheden en beperkingen van juist die agent, niet alleen algemene AI-kennis. De opbouw van deze kennis is proportioneel ingericht naar de rol van de medewerker en het risico van de taken die de agent uitvoert.

Toetsaanpak. De auditor vraagt het instructie- of opleidingsmateriaal per doelgroep op en toetst bij een steekproef van beoordelaars of zij kunnen uitleggen wanneer de uitkomst van deze specifieke agent extra controle behoeft.

Bron: AI-verordening (EU) 2024/1689, artikel 4 (AI-geletterdheid, van kracht sinds 2 februari 2025; door de Digital Omnibus, Verordening (EU) 2026/1744, omgezet in een inspanningsplicht om AI-geletterdheid te bevorderen en te ondersteunen); risicogewicht M; indeling: vertaling van wet; ISO 27001 bijlage A: 6.3 Bewustzijn, opleiding en training

F04 Periodieke verantwoording aan bestuur en auditcommissie

Bestuur en auditcommissie ontvangen periodiek een overzicht van welke agents actief zijn, wat zij hebben gedaan, welke incidenten of afwijkingen zich hebben voorgedaan en welke maatregelen daarop zijn genomen. Dit overzicht staat als vast agendapunt geagendeerd en is geen ad-hoc melding die alleen bij escalatie volgt.

Toetsaanpak. De auditor vraagt de recente verantwoordingsrapportages en vergaderstukken op en stelt vast of agent-gerelateerde risico's en incidenten daadwerkelijk zijn geagendeerd, besproken en van opvolging voorzien.

Bron: Eigen normstelling, aansluitend bij de governance-eisen uit NEN-EN-ISO/IEC 42001:2023 en het IIA AI Auditing Framework (2024); risicogewicht M; indeling: aanscherping; ISO 27001 bijlage A: geen tegenhanger

F05 Verantwoordelijkheidsverdeling bij ingekochte agents

Wanneer een agent door een leverancier wordt geleverd of is ingebed in een ingekochte dienst, is contractueel vastgelegd wie verantwoordelijk is voor het gedrag van de agent, welke zekerheid de leverancier geeft over training, aansturing en grenzen, en hoe de organisatie zelf aan haar verplichtingen blijft voldoen. Dit is schriftelijk vastgelegd, niet stilzwijgend verondersteld.

Toetsaanpak. De auditor vraagt het contract of de verwerkersovereenkomst op bij elke ingekochte of ingebedde agent en toetst of verantwoordelijkheid, aansprakelijkheid en geleverde zekerheden expliciet zijn benoemd.

Afbakening. F05 toetst alleen de contractuele verdeling van verantwoordelijkheid en aansprakelijkheid voor het gedrag van de agent en de zekerheden over aansturing en grenzen; gebruik van gegevens voor training hoort bij F08, subverwerkers en doorgifte bij F09.

Bron: NIST-bevinding februari 2026 (agents draaien doorgaans als generiek serviceaccount zonder eigen identiteit of verantwoording); IMDA Singapore, januari 2026 (verifieerbare digitale identiteit per agent); risicogewicht H; indeling: aanscherping; ISO 27001 bijlage A: 5.19 Informatiebeveiliging in leveranciersrelaties, 5.20 Beveiligingsafspraken in leveranciersovereenkomsten, 5.23 Informatiebeveiliging voor het gebruik van clouddiensten

F06 Recht om de leverancier van de agent te auditen

De organisatie heeft bedongen dat zij, zelf of via een derde, de leverancier van een ingekochte of ingebedde agent kan auditen op relevante aspecten zoals gedrag, logging, toegang en incidentafhandeling. Dit auditrecht is niet alleen contractueel vastgelegd maar ook praktisch uitvoerbaar gebleken.

Toetsaanpak. De auditor controleert of het auditrecht in de overeenkomst is opgenomen en vraagt na of het al eens is uitgeoefend of getoetst, en met welk resultaat.

Bron: Eigen normstelling, aansluitend bij third-party assurance zoals bedoeld in het IIA AI Auditing Framework (2024); risicogewicht M; indeling: al gedekt; ISO 27001 bijlage A: 5.20 Beveiligingsafspraken in leveranciersovereenkomsten, 5.22 Toezicht op en beoordeling van leveranciersdiensten

F07 Aantonen van rechtmatig handelen richting toezichthouder

De organisatie kan op verzoek van een toezichthouder of andere externe partij reconstrueren welke agent, onder welke machtiging, welke handeling heeft verricht, en aantonen dat deze binnen de vastgestelde grenzen bleef. Meldingen en incidenten rond agents zijn vindbaar vastgelegd, inclusief de gegeven opvolging.

Toetsaanpak. De auditor selecteert een afgeronde agent-handeling en vraagt de organisatie deze te reconstrueren zonder hulp van de betrokken ontwikkelaar, en beoordeelt of het incident- en meldingsregister compleet en actueel is.

Afbakening. F07 toetst alleen of de organisatie zonder hulp van de ontwikkelaar een handeling voor een externe partij kan reconstrueren en tegen de vastgestelde grenzen kan leggen, plus het incident- en meldingsregister; eisen aan inhoud, volledigheid en aansluiting van de logging zelf horen bij A06, D01 en D08.

Bron: IMDA Singapore, januari 2026 (spoor van welke agent handelde onder wiens machtiging); AI-verordening (EU) 2024/1689 (toezicht en handhaving); risicogewicht H; indeling: aanscherping; ISO 27001 bijlage A: 5.31 Wettelijke en contractuele eisen, 5.33 Beschermen van registraties, 8.15 Logging

F08 Gebruik van invoer voor training en productverbetering

Voor elke ingekochte of ingebedde agent is schriftelijk vastgelegd of de leverancier de ingevoerde gegevens, de uitvoer of de gebruikstelemetrie gebruikt voor het trainen of verbeteren van zijn modellen of diensten, en zo ja onder welke voorwaarden en met welke mogelijkheid tot uitsluiting. Waar de organisatie gegevens invoert die zij niet mag weggeven, zoals persoonsgegevens, klantgegevens of auditbevindingen, is die uitsluiting aantoonbaar ingesteld en wordt periodiek gecontroleerd dat zij nog actief is, in plaats van eenmalig bij aanschaf verondersteld.

Toetsaanpak. De auditor leest de verwerkersovereenkomst en de productvoorwaarden op dit punt, en controleert vervolgens de feitelijke instelling in de beheeromgeving van de agent. Getoetst wordt of er een periodieke hercontrole is, omdat leveranciers hun voorwaarden en standaardinstellingen tussentijds wijzigen, en of een wijziging daarvan de organisatie bereikt.

Afbakening. F08 toetst alleen of de leverancier invoer, uitvoer of telemetrie gebruikt voor training of productverbetering en of de uitsluiting feitelijk is ingesteld en periodiek wordt gecontroleerd; waar en door wie gegevens verder worden verwerkt hoort bij F09.

Bron: AVG (EU) 2016/679 art. 28; AI-verordening (EU) 2024/1689 (verplichtingen in de aanbiedersketen); eigen normstelling; risicogewicht H; indeling: aanscherping; ISO 27001 bijlage A: 5.20 Beveiligingsafspraken in leveranciersovereenkomsten, 5.22 Toezicht op en beoordeling van leveranciersdiensten, 5.23 Informatiebeveiliging voor het gebruik van clouddiensten, 5.34 Privacy en bescherming van persoonsgegevens

F09 Subverwerkers, doorgifte en rechten van betrokkenen

De organisatie kent de keten achter de agent: welke subverwerkers en welke onderliggende modelaanbieders gegevens verwerken, in welke landen dat gebeurt, en op welke grondslag een doorgifte buiten de EER plaatsvindt. Verzoeken van betrokkenen om inzage, correctie of verwijdering kunnen binnen de wettelijke termijn ook worden uitgevoerd op gegevens die bij die partijen in prompts, logging of geheugen terecht zijn gekomen. Een wijziging in de keten, zoals een nieuwe subverwerker of een verplaatst datacentrum, wordt gemeld en beoordeeld voordat zij van kracht wordt.

Toetsaanpak. De auditor vraagt de actuele subverwerkerslijst van de leverancier op, vergelijkt die met wat in de eigen registers staat, en stelt vast op welke grondslag doorgifte buiten de EER plaatsvindt. Daarnaast wordt één verzoek van een betrokkene gevolgd tot in de agentketen: is het ook uitgevoerd op de gegevens bij de subverwerker, en is dat aantoonbaar. Getoetst wordt tenslotte of de organisatie meldingen van ketenwijzigingen daadwerkelijk ontvangt en beoordeelt.

Afbakening. F09 gaat over de keten buiten de organisatie: subverwerkers en modelaanbieders, landen en doorgiftegrondslag, melding van ketenwijzigingen en het uitvoeren van verzoeken van betrokkenen bij die partijen; opslagplaatsen en bewaartermijnen in de eigen agentomgeving horen bij D09, trainingsgebruik bij F08 en de algemene contractuele verantwoordelijkheid bij F05.

Bron: AVG (EU) 2016/679 art. 28 lid 2-4, hoofdstuk V en art. 12-22; aansluitend bij ISO/IEC 27001:2022 bijlage A 5.14, 5.22 en 5.23 (niet-letterlijk overgenomen); risicogewicht H; indeling: vertaling van wet; ISO 27001 bijlage A: 5.20 Beveiligingsafspraken in leveranciersovereenkomsten, 5.21 Informatiebeveiliging in de ICT-toeleveringsketen, 5.22 Toezicht op en beoordeling van leveranciersdiensten, 5.34 Privacy en bescherming van persoonsgegevens

Vervallen: B07 (maximale ketendiepte en externe diensten) is opgegaan in A08; het nummer wordt niet hergebruikt.