Samen aanjagen van vernieuwing

Handreiking AI en Informatiebeveiliging binnen onderwijsinstellingen

Voor veilig en verantwoord AI-gebruik

Auteur(s): HBO CISO Beraard – Werkgroep AI
Matijs van Hilst, Kees Codée, Wilbert Hepping, Joop Selen

Versie: 1.0
Datum: 12 november 2025
Kenmerk: Handreiking AI en Informatiebeveiliging

Documentinformatie

Dit document is een op zichzelf staand hulpmiddel dat ondersteuning biedt bij het verantwoord gebruik van AI binnen onderwijsinstellingen. Het biedt handvatten voor risicobeheersing en volwassenheidsniveaus met behulp van bestaande AI-kaders en het SURFaudit Toetsingskader IB.

Versiebeheer

Versie	Datum	Auteur	Verwerking
0.1	24-2-2025	Matijs van Hilst Kees Codée Wilbert Hepping Joop Selen	Opzet
1.0	12-12-2025	Ed de Vries	Definitieve versie

Distributielijst

Versie	Datum	Ontvanger	Doel

Vaststelling

Versie	Datum	Vastgesteld door	Vastgesteld op

Samenhang met andere documenten

Naam	Bovenliggend	Gelijk niveau	Onderliggend

Verwijzingen naar SURFaudit Toetsingskader en ISO27001

Kader	Verwijzing (tags)
SURFaudit Toetsingskader	Alle domeinen

Inhoudsopgave

Samenvatting

Deze handreiking biedt instellingen inzicht in de risico's van AI en hoe deze effectief beheerst kunnen worden met behulp van bestaande AI-kaders en het SURFaudit Toetsingskader IB. Het koppelt AI-risico's aan relevante domeinen van het toetsingskader IB en geeft aanbevelingen voor gewenste volwassenheidsniveaus per domein. Hiermee helpt het instellingen om AI veilig en verantwoord te implementeren zonder nieuwe kaders te hoeven ontwikkelen.

1 Inleiding

In de huidige digitale samenleving is Artificial Intelligence (AI) niet meer weg te denken. AI-technologieën worden steeds vaker toegepast binnen instellingen om processen te optimaliseren, besluitvorming te ondersteunen en innovatieve diensten te leveren. Hoewel AI tal van voordelen biedt, brengt het ook nieuwe risico's en uitdagingen met zich mee op het gebied van informatiebeveiliging. Denk hierbij aan dataprivacy, integriteit van beslissingsprocessen en de kwetsbaarheid voor manipulatie of aanvallen op AI-modellen.

Dit document is opgesteld namens de werkgroep AI van het HBO CISO Beraad als hulpmiddel voor instellingen om effectief om te gaan met de informatiebeveiligingsrisico's van AI. Het doel is om inzicht te geven in de relatie tussen AI en het SURFaudit Toetsingskader IB voor informatiebeveiliging, zonder daarbij een compleet nieuw kader te creëren. In plaats daarvan wordt er gekeken hoe de bestaande beheersdoelstellingen uit het SURFaudit Toetsingskader IB toegepast kunnen worden op AI-gerelateerde risico's.

Doelgroep en gebruik van dit document

Dit document is bedoeld voor CISO's, IT-beveiligingsprofessionals, AI-specialisten en beleidsmakers binnen instellingen die te maken hebben met de implementatie en beveiliging van AI-systemen. Het biedt een praktisch kader om AI-risico's te identificeren, te beoordelen en te beheersen binnen de bestaande structuren van het SURFaudit Toetsingskader IB.

Met dit document willen we instellingen ondersteunen bij het verantwoord en veilig inzetten van AI-technologieën, waarbij informatiebeveiliging op een gepaste en effectieve manier is geborgd.

2 Bestaande AI-kaders

In het snel veranderende landschap van Artificial Intelligence (AI) zijn er tal van kaders en richtlijnen ontwikkeld om instellingen te ondersteunen bij het verantwoord en veilig inzetten van AI-technologieën. Deze kaders bieden houvast voor het beheersen van AI-gerelateerde risico's, het waarborgen van ethiek en transparantie, en het naleven van wet- en regelgeving.

Dit hoofdstuk geeft een overzicht van de meest relevante standaarden, kaders en richtlijnen die momenteel beschikbaar zijn. Hierbij wordt kort toegelicht hoe deze kaders instellingen kunnen helpen bij het beheersen van risico's, het waarborgen van ethiek en transparantie, en het voldoen aan wettelijke eisen. Door inzicht te bieden in deze kaders kunnen instellingen beter geïnformeerde keuzes maken bij het implementeren en beveiligen van AI-systemen.

2.1 ISO/IEC 42001:2023

De ISO/IEC 42001:2023[1] is een internationale standaard die richtlijnen biedt voor het opzetten van een AI-managementsysteem (AIMS) binnen instellingen. Deze standaard helpt bij het ontwikkelen van een raamwerk voor AI-governance, met nadruk op ethiek, transparantie en risicobeheer. Het biedt een gestructureerde aanpak voor het beveiligen van AI-systemen, inclusief richtlijnen voor risicobeoordeling, incidentrespons en compliance met internationale regelgeving.

Door te voldoen aan deze standaard kunnen instellingen aantonen dat hun AI-toepassingen betrouwbaar zijn en in lijn zijn met internationale best practices. Bovendien helpt deze standaard instellingen bij het creëren van transparante en uitlegbare AI-modellen, wat bijdraagt aan vertrouwen en verantwoording in AI-besluitvormingsprocessen.

2.2 ISO/IEC 23894:2023

De ISO/IEC 23894:2023 [2]is een internationale standaard die richtlijnen biedt voor het beheren van AI-veiligheidsrisico’s binnen instellingen. De stanaard richt zich specifiek op het identificeren, evalueren en mitigeren van risico’s die voortkomen uit het gebruik van AI-systemen. Hierbij wordt niet alleen gekeken naar de technische risico’s, zoals modelmanipulatie en datalekken, maar ook naar de etische risico’s zoals bias en discriminatie.
De standaard help instellingen om een gestructureerde aanpak te ontwikkelen voor AI-risicobeheer, met nadruk op transparatie, verantwoordelijkheid en ethiek.
ISO/IEC 23894:2023 ondersteunt instellingen bij het integreren van AI-veiligheidsmaatregelen in hun bestaande informatiebeveiligingsbeleid, waardoor een holistische benadering van AI-risicobeheer mogelijk wordt. Hierdoor kunnen instellingen AI op een verantwoorde en veilige manier inzetten, met respect voor ethische normen en wettelijke verplichtingen.

[1] https://www.iso.org/standard/81230.html

[2] https://www.iso.org/standard/77304.html

In verhouding tot ISO 42001
ISO/IEC 42001 richt zich op het opzetten van een AI-managementsysteem (AIMS) binnen een organisatie, terwijl ISO/IEC 23894 zich specifiek concentreert op het risicomanagementaspect van AI.

2.3 Algoritmekader

Het Algoritmekader[3] van het Nederlandse Ministerie van Binnenlandse Zaken biedt specifieke richtlijnen voor het transparant en verantwoord inzetten van algoritmen, met name binnen de publieke sector. Het kader richt zich op transparantie, verantwoording en ethiek en helpt instellingen om algoritmen op een eerlijke en uitlegbare manier te gebruiken.

Dit kader ondersteunt instellingen bij het waarborgen van transparantie in AI-besluitvormingsprocessen, wat belangrijk is voor het voldoen aan wetgeving zoals de AVG. Daarnaast biedt het handvatten voor het uitvoeren van audits en het identificeren van ethische risico's. Het kader helpt bij het ontwikkelen van strategieën om algoritmen te beschermen tegen manipulatie en aanvallen, zoals adversarial attacks, en bevordert daarmee de integriteit van AI-systemen.

2.4 Ethische Richtlijnen voor Betrouwbare AI

De Ethische Richtlijnen voor Betrouwbare AI[4] van de Europese Commissie zijn ontworpen om ervoor te zorgen dat AI-systemen betrouwbaar, transparant en ethisch verantwoord zijn. Deze richtlijnen omvatten zeven kernvereisten, waaronder menselijke controle, privacy, transparantie en maatschappelijke welzijn.

Instellingen kunnen deze richtlijnen gebruiken om ethische risico's te identificeren en te beheren, zoals bias in AI-modellen. Door de focus op privacy ondersteunen deze richtlijnen instellingen bij het naleven van de AVG en andere privacywetgevingen. Daarnaast helpen ze bij het opzetten van audittrails en transparante besluitvormingsprocessen, wat bijdraagt aan verantwoording en maatschappelijke acceptatie van AI-systemen.

2.5 AI Act

De AI Act[5] is een Europese wet die AI-toepassingen classificeert op basis van hun risiconiveau (van laag tot hoog risico) en verplichtingen oplegt aan zowel ontwikkelaars als gebruikers van AI-systemen. Deze wetgeving biedt een gemeenschappelijk regelgevingskader binnen de EU voor AI en stelt duidelijke eisen aan transparantie, veiligheid en ethiek.

Instellingen kunnen deze wet gebruiken om AI-toepassingen te classificeren op basis van risiconiveau en gerichte beveiligingsmaatregelen te implementeren. De AI Act helpt bij het waarborgen van naleving van Europese regelgeving, wat cruciaal is voor instellingen die AI in de EU gebruiken. Daarnaast biedt de wetgeving richtlijnen voor beveiliging en incidentrespons, wat helpt bij het ontwikkelen van incidentresponsplannen specifiek voor AI.

[3] https://minbzk.github.io/Algoritmekader/

[4] https://ec.europa.eu/commission/presscorner/detail/nl/ip_19_1893

[5] https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng?eliuri=eli%3Areg%3A2024%3A1689%3Aoj&locale=nl

2.6 UNESCO richtlijnen en aanbevelingen voor AI

UNESCO heeft meerdere richtlijnen en aanbevelingen ontwikkeld om ethisch en verantwoord gebruik van AI te bevorderen. Deze kaders richten zich op mensenrechten, inclusiviteit, duurzaamheid en digitale geletterdheid, met een bijzondere focus op het onderwijs. Hieronder volgt een overzicht in tabelvorm van de belangrijkste richtlijnen en aanbevelingen, inclusief hun doelen en toepassingsgebieden.

Richtlijn / aanbeveling	Jaar	Doel	Toepassingsgebied
Aanbeveling over de Ethiek van AI[6]	2021	Bevorderen van ethisch AI-gebruik met nadruk op mensenrechten, inclusiviteit en duurzaamheid.	Instellingen in alle sectoren die ethisch AI willen toepassen.
Richtlijn voor Beleidsmedewerkers over AI in Onderwijs[7]	2021	Verantwoord en transparant gebruik van AI in onderwijs- en leeromgevingen.	Beleidsmakers en onderwijsinstellingen.
Richtlijn voor Generatieve AI in Onderwijs en Onderzoek[8]	2023	Verantwoord gebruik van generatieve AI met aandacht voor ethiek, auteursrechten en transparantie.	Onderwijsinstellingen en academische onderzoeksinstituten.
AI-Vaardigheden voor Docenten en Studenten[9]	2023	Bevorderen van AI-geletterdheid en ethisch bewustzijn bij docenten en studenten.	Onderwijsinstellingen voor alle niveaus.

2.7 AI Ethiek volwassenheidsmodel

Het 'AI Ethiek volwassenheidsmodel[10]' is een holistisch raamwerk dat helpt bij het implementeren van ethische AI-praktijken binnen instellingen. Het model omvat zes dimensies die essentieel zijn voor het operationaliseren van AI-ethiek: Awareness en Cultuur, Beleid, Governance, Communicatie en Training, Ontwikkelingsprocessen en Tooling. Elk van deze dimensies kent vijf volwassenheidsniveaus (ad-hod tot geoptimaliseerd). Het model help instellingen te evalueren hoever zij gevorderd zijn in de operationalisatie van AI-ethiek en biedt een groeipad naar een hoger volwassenheidsniveau.

2.8 Conclusie

De zeven besproken AI-kaders en richtlijnen bieden een solide basis voor het verantwoord en ethisch inzetten van AI binnen onderwijsinstellingen. Elk van deze kaders heeft een eigen focusgebied, variërend van transparantie en ethiek tot privacy en inclusiviteit. Ze helpen onderwijsinstellingen om AI op een eerlijke, transparante en verantwoorde manier te

[6] https://www.unesco.nl/nl/artikel/unesco-commissie-brengt-nederlandse-brochure-uit-over-kunstmatige-intelligentie

[7] https://unesdoc.unesco.org/ark:/48223/pf0000376709

[8] https://unesdoc.unesco.org/ark:/48223/pf0000386693

[9] https://www.unesco.nl/nl/artikel/welke-ai-vaardigheden-moeten-docenten-en-leerlingen-hebben

[10] https://link.springer.com/article/10.1007/s43681-022-00228-7

implementeren, met speciale aandacht voor ethische risico's zoals bias, privacy-inbreuken en discriminatie. Daarnaast stimuleren ze digitale geletterdheid en het verantwoord gebruik van AI in onderwijsomgevingen.

Hoewel deze kaders waardevolle richtlijnen bieden, wordt de praktijk vaak gekenmerkt door een overvloed aan verschillende standaarden, kaders en aanbevelingen. Dit maakt het voor onderwijsinstellingen uitdagend om een duidelijke en toepasbare strategie te ontwikkelen. Het gebrek aan uniformiteit leidt tot fragmentatie, waarbij instellingen geconfronteerd worden met de vraag welke richtlijnen gevolgd moeten worden en hoe deze effectief in de bestaande structuren geïntegreerd kunnen worden.

De komst van de AI Act van de Europese Unie voegt een extra laag complexiteit toe, aangezien deze wetgeving juridische verplichtingen introduceert die onderwijsinstellingen moeten naleven. In tegenstelling tot de overige richtlijnen is de AI Act wettelijk bindend, wat betekent dat onderwijsinstellingen concrete stappen moeten ondernemen om aan deze wetgeving te voldoen.

Juist vanwege deze complexiteit is dit document tot stand gekomen. Het doel is om onderwijsinstellingen te helpen AI te mappen op een bestaand kader dat zij al gebruiken, het SURFaudit Toetsingskader IB. Door AI-risico's en beheersdoelstellingen te koppelen aan een bekend en bewezen raamwerk, kunnen onderwijsinstellingen effectiever inspelen op de uitdagingen van AI zonder verstrikt te raken in een veelheid aan externe kaders.

Op deze manier biedt dit document niet alleen een overzicht van relevante AI-richtlijnen, maar ook een praktische aanpak voor onderwijsinstellingen om AI op een veilige, transparante en ethisch verantwoorde manier te integreren binnen hun bestaande informatiebeveiligingsstructuren.

3 Beoogde volwassenheidsniveaus per domein

Dit hoofdstuk biedt een handreiking voor de gewenste volwassenheidsniveaus per SURFaudit Toetsingskader IB-domein om AI op een veilige en verantwoorde manier binnen de organisatie te kunnen inzetten. In lijn met het SURFAudit Toetsingskader IB wordt gebruik gemaakt van de niveaus 1 tot en met 5, waarbij niveau 1 een ad-hoc aanpak vertegenwoordigt en niveau 5 staat voor een geoptimaliseerde en volledig geïntegreerde benadering van informatiebeveiliging.

De implementatie van AI brengt specifieke risico's met zich mee op het gebied van ethiek, privacy, transparantie en beveiliging. Om deze risico's effectief te beheersen, is het noodzakelijk om per domein een passend volwassenheidsniveau te bereiken. Het gewenste niveau kan echter per organisatie verschillen, afhankelijk van de aard van de AI-toepassingen, de gevoeligheid van de verwerkte data en het risicoprofiel van de organisatie.

De voorgestelde plotting van risico's en beoogde niveaus in dit hoofdstuk is bedoeld als een opzet en dient als leidraad om risico's zo klein mogelijk te maken. Afhankelijk van hun eigen context en risicobereidheid kunnen instellingen de gewenste niveaus aanpassen aan hun specifieke situatie. Deze benadering biedt een pragmatische richtlijn voor het behalen van een volwassen en veilig AI-beleid, zonder dat één vaste standaard wordt opgelegd.

Domein	**ID**	**Categorie beheersmaatregel**	**Gewenst niveau**
Governance	GO.01	Strategie	**4**
	GO.02	Beleid	**4**
	GO.03	Planning / Roadmap	**4**
	GO.04	Architectuur	**3**
	GO.05	Onafhankelijke toetsing	**4**
Organisatie	OR.01	Eigenaarschap, rollen, verantwoording en verantwoordelijkheid	**4**
	OR.02	Functiescheiding	**4**
Risicobeheer	RM.01	Informatie risico- management kader	**4**
	RM.02	Risicobeoordeling	**4**
	RM.03	Plan voor behandeling en beperking van risico's (inclusief risicoacceptatie)	**4**
Personeelsbeheer	HR.01	Werving	**3**
	HR.02	Certificering, training en scholing	**4**
	HR.03	Afhankelijkheid van individuen	**3**
	HR.04	Verandering of beëindiging van functie	**3**
	HR.05	Kennisdeling	**3**
	HR.06	Veiligheidsbewustzijn	**4**
Configuratiebeheer	CO.01	Identificatie en onderhoud van configuratie-items	**3**
	CO.02	Configuratiedatabase en baseline	**3**
Incident/probleembeheer	IM.01	incident management	**3**
	IM.02	Incident escalatie	**3**
	IM.03	Incidentrespons op (cyber) beveiligingsincidenten	**4**

	IM.04	Probleembeheer	**3**
Wijzigingsbeheer	CH.01	Normen en procedures voor aanpassingen	**3**
	CH.02	Impact assessment, prioriteren en autoriseren	**3**
	CH.03	Noodaanpassingen	**3**
	CH.04	Testomgeving	**3**
	CH.05	Testen van aanpassingen	**3**
	CH.06	Promotie naar productie	**3**
Systeemontwikkeling	SD.01	Methodiek voor veilige softwareontwikkeling en -implementatie	**3**
	SD.02	Toegang tot de productieomgeving door ontwikkelaars	**4**
	SD.03	Data conversie en/of migratie	**3**
Gegevensbeheer	DM.01	Data (en systeem) eigenaarschap	**4**
	DM.02	Classificatie	**4**
	DM.03	Beveiligingseisen voor gegevensbeheer	**4**
	DM.04	Inrichting van opslag en retentie	**4**
	DM.05	Uitwisseling van (gevoelige) gegevens	**4**
	DM.06	Verwijderen van data	**4**
Identiteits- en toegangsbeheer	ID.01	Toegangsrechten	**4**
	ID.02	Administratie van toegangsrechten	**4**
	ID.03	Super Users	**3**
	ID.04	Noodtoegang (envelopprocedure/breek-het-glasprocedure)	**3**
	ID.05	Periodieke beoordeling van toegangsrechten	**4**
Beveiligingsbeheer	SM.01	Security baselines	**3**
	SM.02	Authenticatiemechanismes	**4**
	SM.03	Mobiele apparaten en telewerken	**3**
	SM.04	Logging	**4**
	SM.05	Testen van, inspectie van en toezicht op beveiliging	**4**
	SM.06	Patchbeheer	**3**
	SM.07	Beheer van bedreigingen en kwetsbaarheden	**3**
	SM.08	Beschikbaarheid en bescherming van infrastructuur	**3**
	SM.09	Onderhoud van de infrastructuur	**3**
	SM.10	Beheer van cryptografische sleutels	**3**
	SM.11	Network security	**3**
	SM.12	Beheersing van malware-aanvallen	**3**
	SM.13	Bescherming van beveiligingstechnologie	**3**
Fysieke beveiliging	PH.01	Fysieke beveiligingsmaatregelen	**3**
	PH.02	Beheer van fysieke toegangsrechten	**3**

IT-operatie	OP.01	Job processing	**3**
	OP.02	Procedures voor back-up en herstel	**3**
	OP.03	Capaciteits- en prestatiebeheer	**3**
Bedrijfscontinuïteitsbeheer	BC.01	Bedrijfscontinuïteits-planning	**3**
	BC.02	Testen van Disaster recovery	**3**
	BC.03	Externe back-upopslag	**3**
	BC.04	Gegevensreplicatie	**3**
	BC.05	Crisisbeheer	**3**
Ketenbeheer	SC.01	Service level overeenkomst	**4**
	SC.02	Service level beheer	**4**
	SC.03	Risicobeheer van leveranciers	**4**
	SC.04	Interne beheersing bij derden	**4**

4 Beheersmaatregelen per domein

Dit hoofdstuk beschrijft hoe AI gerelateerde risico's en beheersmaatregelen gekoppeld kunnen worden aan de domeinen uit het SURFaudit Toetsingskader IB. Maatregelen zijn veelal gekoppeld aan controls uit de ISO/IEC 42001:2023, en uitgebreid met een niveau passend bij de niveau's van het SURFaudit Toetsingskader IB.
Het doel is om inzicht te geven in de impact van AI op elk domein en om instellingen te helpen effectieve maatregelen te implementeren die aansluiten bij bestaande beheersdoelstellingen. Dit zorgt voor een gestructureerde aanpak van AI-risico's zonder een compleet nieuw kader te moeten ontwikkelen.

4.1 Governance

AI introduceert nieuwe ethische en juridische vraagstukken die een duidelijke governance-structuur vereisen. Instellingen moeten beleid ontwikkelen voor het verantwoord gebruik van AI, inclusief ethische richtlijnen en transparantie-eisen. Governance omvat tevens het vaststellen van rollen en verantwoordelijkheden voor AI-beheer en -toezicht.

Effectieve maatregelen:

Maatregel	Niveau	ISO42001
Integreer AI-governance in bestaande strategie en beleid.	3	A.2.2 A.2.3
Zorg voor een audit-trail en verantwoording van AI-besluitvormingsprocessen.	4	A.2.4

4.2 Organisatie

Bij de inzet van AI moeten instellingen zorgen voor een heldere structuur met duidelijke rollen en verantwoordelijkheden. Het organiseren van multidisciplinaire teams met AI-experts, ethici en IT-beveiligingsspecialisten is essentieel voor een veilige en verantwoorde implementatie.

Effectieve maatregelen:

Maatregel	Niveau	ISO42001
Stel een multidisciplinaire AI-commissie in voor toezicht en verantwoording. De commissie moet bestaand uit vertegenwoordigers vanuit IT, privacy, juridische zaken en ethiek.	3	A.3.2
Definieer rollen en verantwoordelijkheden voor AI-governance.	3	A.3.2
Zorg voor periodieke training en bewustwording van AI-risico's.	3	A.4.6
Definieer een proces waarbij zorgen over de rol van de organisatie m.b.t. AI kunnen worden gerapporteerd.	4	A.3.3

4.3 Risicobeheer

AI introduceert nieuwe risico's zoals bias, modelmanipulatie, privacy-inbreuken en datalekken. Traditionele risicobeheersingsmethoden moeten worden aangepast om AI-specifieke risico's te identificeren, evalueren en te mitigeren. Dit is des te relevanter nu de AI Act van de Europese Unie juridische verplichtingen introduceert die instellingen verplichten om AI-risico's te classificeren op basis van risiconiveaus (laag, beperkt, hoog en onaanvaardbaar risico).

De AI Act vereist dat instellingen een risicobeoordeling uitvoeren voor AI-systemen en voorziet in strengere vereisten voor hoog-risico AI-toepassingen, zoals transparantie, robuustheid en nauwkeurigheid. Instellingen moeten aantonen dat zij passende risicobeheersmaatregelen hebben geïmplementeerd en dat AI-systemen voldoen aan ethische en wettelijke normen.

Effectieve maatregelen:

Maatregel	Niveau	ISO42001
Integreer AI-risicobeoordelingen in bestaande risicobeheerprocessen en leg de resultaten van deze beoordelingen vast.	3	A.5.2
Zorg voor compliance aan wet- en regelgeving, waaronder de AI Act	3	A.5.3
Classificeer AI-systemen op basis van risiconiveaus zoals gedefinieerd in de AI Act. *(Indien beschikbaar kunnen Classificaties worden overgenomen vanuit SURF Vendor Compliance).*	3	A.5.2
Voer periodieke modelaudits uit om bias, discriminatie en andere ethische risico's te detecteren	4	A.5.4 A.5.5
Beoordeel periodiek de impact van AI op bestaande bedrijfsprocessen en zorg dat risicobeheersing aansluit op de wettelijke vereisten van de AI Act	4	A.5.4 A.5.5
Implementeer een Algoritme-register waarin bovenstaande wordt vastgelegd[11]	3	A.5

4.4 Personeelsbeheer

AI verandert de vaardigheden die nodig zijn binnen instellingen. Medewerkers moeten worden getraind in AI-ethiek, gegevensbeheer en beveiligingsprotocollen. Daarnaast moet bewustwording worden gecreëerd over de risico's van AI-manipulatie en datalekken.

Effectieve maatregelen:

Maatregel	Niveau	ISO42001
Versterk AI-competenties door gerichte opleidingen en bewustwordingsprogramma's.	3	A.4.6

[11] Dit statement is een kopie van RB.04 VWN3 a uit het SRUF Toetsingskader Privacy.

Integreer AI binnen het bestaande security-awareness programma.	3	A4.6
Stel ethische richtlijnen op voor AI-gebruik binnen de organisatie.	3	A.9.2 A.9.3
Beoordeel periodiek of medewerkers beschikken over de juiste competenties voor hun rol binnen AI-systemen.	4	A.4.6

4.5 Configuratiebeheer

AI-modellen en algoritmen moeten goed worden geconfigureerd en beheerd om beveiligingsrisico's te minimaliseren. Versiebeheer en controle op wijzigingen zijn cruciaal om de integriteit van AI-systemen te waarborgen.

Effectieve maatregelen:

Maatregel	Niveau	ISO42001
Leg AI-componenten (modellen, datasets, scripts, pipelines) vast in de bestaande configuratiebeheer-database (CMDB).	3	A.4.2 t/m A.4.6
Documenteer wijzigingen in AI-configuraties (versies van modellen, parameters, datasets) ten behoeve van transparantie en traceerbaarheid.	3	A.6.2.3 A.6.2.8
Evalueer periodiek de configuratie-instellingen op beveiligingsrisico’s.	4	
Identificeer en documenteer alle middelen (kennis, data, tooling, componenenten) die nodig zijn voor het managen van AI gedurende haar hele levenscyclus (ISO42001 A4)	4	A.4.2 t/m A.4.6

4.6 Incident- en Probleembeheer

Incidenten met AI kunnen impact hebben op besluitvorming, privacy en veiligheid. In veel gevallen kunnen AI-incidenten echter binnen bestaande incident- en probleembeheerprocessen worden afgehandeld, mits deze processen flexibel genoeg zijn om AI-specifieke afwijkingen te herkennen en te beheren.

Effectieve maatregelen:

Maatregel	Niveau	ISO42001
Breid het bestaande incidentbeheerproces uit met herkenning en behandeling van AI-specifieke incidenttypen.	3	A.6.2.8 A.8.4

4.7 Wijzigingsbeheer

AI-systemen vereisen frequente updates en aanpassingen. Wijzigingen kunnen invloed hebben op de prestaties en veiligheid van AI-modellen. In veel gevallen kan AI-wijzigingsbeheer binnen bestaande wijzigingsprocessen worden geïntegreerd maar zal dit meer tijd vereisen dan andere applicaties.

Effectieve maatregelen:

Maatregel	Niveau	ISO42001
Behandel wijzigingen in AI-modellen binnen het bestaande wijzigingsbeheerproces.	3	A.6.2.5 A.6.2.6
Voer (geautomatiseerde) impactanalyses uit bij wijzigingen in AI-toepassingen en AI-modellen.	4	A.6.2.4

4.8 Systeemontwikkeling

AI vereist specifieke ontwikkelmethoden om veiligheid, ethiek en robuustheid te waarborgen. Hierbij kunnen veel bestaande beheersmaatregelen voor systeemontwikkeling toegepast worden, mits deze worden uitgebreid met AI-specifieke overwegingen zoals ethische risico's, transparantie en modelintegriteit. Het SURFaudit Toetsingskader IB-domein voor systeemontwikkeling richt zich op het gestructureerd en veilig ontwikkelen en beheren van software, wat cruciaal is voor AI-modellen en algoritmen.

Bestaande maatregelen zoals Security-by-Design, Privacy-by-Design en Privacy-by-Default zijn ook relevant voor AI-ontwikkelingen, vooral gezien de eisen die de AI Act stelt aan transparantie, ethiek en risicobeheersing voor hoog-risico AI-toepassingen.

Effectieve maatregelen:

Maatregel	Niveau	ISO42001
Definieer duidelijke doelstellingen voor een verantwoorde ontwikkeling van AI modellen en/of AI systemen.	3	A.6.1.2 A.6.1.3
Integreer ethische en privacy-impact assessments in het ontwikkelproces voor AI, zoals vereist door de AI Act.	3	
Zorg voor transparante documentatie van AI-modellen en algoritmen, zodat auditability en verantwoording gewaarborgd zijn.	4	A.6.2.3 A.6.2.7
Controleer of bestaande rollen- en toegangsbeheerprocedures voldoende zijn voor AI-systemen, met name om schrijftoegang tot productieomgevingen te beperken en modelintegriteit te waarborgen.	4	A.6.2.4

4.9 Gegevensbeheer

AI is sterk afhankelijk van data, wat specifieke eisen stelt aan gegevensbeheer en privacy. Binnen het SURFaudit Toetsingskader IB-domein voor gegevensbeheer zijn er uitgebreide beheersmaatregelen die gericht zijn op het beschermen van informatiemiddelen, het classificeren van data en het waarborgen van integriteit en vertrouwelijkheid. Veel van deze maatregelen zijn direct toepasbaar op AI-systemen, mits er extra aandacht wordt besteed aan de specifieke risico's en eisen die AI met zich meebrengt, zoals transparantie, bias en ethische data-analyse.

De AI Act vereist daarnaast dat instellingen transparantie en verantwoording kunnen aantonen voor de data die wordt gebruikt in AI-modellen, met name voor hoog-risico AI-toepassingen. Dit betekent dat gegevensbeheer voor AI striktere controles moet omvatten op dataclassificatie, opslag, uitwisseling en verwijdering.

Effectieve maatregelen:

Maatregel	Niveau	ISO42001
Pas bestaande dataclassificatieschema's toe op alle datasets die worden gebruikt voor AI-training en -testen om de juiste beveiligingsmaatregelen te waarborgen en te voldoen aan de eisen van de AI Act.	3	A.7.1 t/m A.7.6
Zorg dat procedures voor dataopslag, archivering, uitwisseling en verwijdering ook van toepassing zijn op AI-datasets, zodat transparantie, traceerbaarheid en naleving van wetgeving gewaarborgd zijn.	4	

4.10 Identiteits- en Toegangsbeheer

Identiteits- en toegangsbeheer is cruciaal voor het waarborgen van de vertrouwelijkheid, integriteit en beschikbaarheid van gegevens in AI-systemen. Binnen het SURFAudit Toetsingskader IB richt dit domein zich op het beheer van gebruikersrechten, authenticatie, autorisatie en het traceren van gebruikersactiviteiten. AI-systemen vereisen dezelfde strikte beveiliging als andere informatiesystemen, maar brengen specifieke risico's met zich mee, zoals ongeautoriseerde toegang tot gevoelige datasets of manipulatie van AI-modellen.

Effectieve maatregelen:

Maatregel	Niveau	ISO42001
Neem AI op in de procedures voor het beheer van toegangsrechten.	3	A.9.4
Zorg voor (geautomatiseerde) identificatie, authenticatie en autorisatie van gebruikers die AI-modellen trainen, testen en implementeren.	3	A.6.2.8
Implementeer logging en monitoring, zodat toegang tot AI-data en -modellen volledig traceerbaar is.	4	A.6.2.8 A.9.4

Geef een AI-toepassing pas toegang tot data nadat deze data is geclassificeerd, de toegangsrechten zijn vastgelegd en vastgesteld is dat de AI-toepassing deze rechten herkent en respecteert.	3	A.7.4

4.11 Beveiligingsbeheer

Beveiligingsbeheer richt zich op het waarborgen van de vertrouwelijkheid, integriteit en beschikbaarheid van informatie binnen de organisatie. In de context van AI is het belangrijk om niet alleen de beveiliging van gegevens te waarborgen, maar ook de integriteit van AI-modellen en algoritmen. Het SURFAudit Toetsingskader IB stelt dat beveiligingsbaselines formeel moeten zijn vastgesteld, goedgekeurd door senior management en periodiek gecontroleerd op naleving. Voor AI betekent dit dat beveiligingsmaatregelen moeten worden uitgebreid naar gegevens die worden gebruikt voor het trainen en testen van AI-modellen, alsook naar de modellen zelf.

In veel gevallen kunnen AI-specifieke risico's worden beheerst door de bestaande beveiligingsmaatregelen uit het SURFaudit Toetsingskader IB te versterken en te combineren met AI-specifieke controles op modelintegriteit en gegevensbescherming. Het is belangrijk om te evalueren of de huidige beveiligingsmaatregelen voldoen aan de eisen van de AI Act, vooral voor hoog-risico AI-toepassingen. Hierbij moet specifiek aandacht worden besteed aan modelmanipulatie, bias, en de betrouwbaarheid van AI-uitkomsten.

Effectieve maatregelen:

Maatregel	Niveau	ISO42001
Evalueer periodiek of de huidige beveiligingsmaatregelen voldoen aan de eisen van de AI Act, met speciale aandacht voor modelmanipulatie, bias en de betrouwbaarheid van AI-uitkomsten, met name voor hoog-risico AI-toepassingen en AI-toepassingen welke gebruik maken van gevoelige data.	4	A.6.2.6 A.6.2.8 A.7.4

4.12 Fysieke beveiliging

Fysieke beveiliging richt zich op het beschermen van fysieke locaties, apparatuur en gegevens tegen ongeautoriseerde toegang, schade en diefstal. In de context van AI is het belangrijk om niet alleen de toegang tot servers en datacenters te beveiligen, maar ook de fysieke opslag van AI-modellen en trainingsdata. Het SURFaudit Toetsingskader IB vereist dat fysieke beveiligingsmaatregelen gebaseerd zijn op een risico-analyse, goedgekeurd door senior management en gedocumenteerd in beleid en procedures.

De bestaande maatregelen uit het SURFaudit Toetsingskader IB voor fysieke beveiliging zijn over het algemeen voldoende om ook AI-specifieke risico's te beheersen. Denk hierbij aan toegangscontrole voor serverruimtes waar AI-modellen worden gehost en het beveiligen van dataopslaglocaties. Er zijn geen aanvullende AI-specifieke maatregelen nodig, zolang de bestaande beveiligingsmaatregelen consequent worden toegepast en up-to-date blijven.

4.13 IT-operatie

De bestaande maatregelen uit het SURFaudit Toetsingskader IB voor IT-operatie zijn over het algemeen voldoende om ook AI-specifieke risico's te beheersen. Denk hierbij aan job scheduling voor AI-training en inferentieprocessen, back-ups van datasets en modellen, en prestatiemonitoring van AI-infrastructuur. Er zijn geen aanvullende AI-specifieke maatregelen nodig, zolang de bestaande maatregelen consequent worden toegepast en afgestemd zijn op de specifieke eisen van AI-workloads.

4.14 Bedrijfscontinuïteitsbeheer

De bestaande maatregelen uit het SURFaudit Toetsingskader IB voor bedrijfscontinuïteitsbeheer zijn over het algemeen voldoende om ook AI-specifieke risico's te beheersen. Dit omvat het opnemen van AI-modellen en datasets in back-up- en herstelplannen. Het is belangrijk dat AI-systemen worden meegenomen in de bedrijfsimpactanalyse en dat herstelplannen hierop worden afgestemd.

Effectieve maatregelen:

Maatregel	Niveau	ISO42001
Neem AI-modellen, datasets en pipelines expliciet op in bestaande back-up- en herstelprocedures	3	
Zorg dat AI-systemen onderdeel zijn van de bedrijfsimpactanalyse en herstelstrategieën (inclusief testscenario's).	3	

4.15 Ketenbeheer

De AI Act stelt aanvullende eisen aan ketenbeheer, met name voor hoog-risico AI-systemen. Instellingen moeten kunnen aantonen dat leveranciers van AI-componenten voldoen aan de eisen voor transparantie, betrouwbaarheid en ethiek. Dit omvat het beoordelen van de betrouwbaarheid van leveranciers, het waarborgen van de naleving van data privacy-regelgeving en het vastleggen van duidelijke afspraken in contracten en SLA's. Deels denken we dat dit ook belegd kan worden bij Surf.

Effectieve maatregelen:

Maatregel	Niveau	ISO42001
Specificeer en documenteer vereisten voor nieuwe AI systemen of AI-uitbreidingen bij bestaande systemen.	3	A.6.2.2
Leg afspraken vast in SLA's over transparantie, betrouwbaarheid en ethiek voor alle externe AI-componenten, in lijn met de eisen van de AI Act.	3	A.10.2 t/m A.10.4

Implementeer leveranciersrisicobeheer dat rekening houdt met de betrouwbaarheid van AI-leveranciers, compliance met data privacy en beveiligingseisen, en mogelijke ethische risico's.	3	A.5.2 A.10.3
Evalueer periodiek de naleving van contractuele verplichtingen en compliance met de AI Act door externe dienstverleners van AI-modellen, datasets en cloud-diensten.	4	A.10.2