

AI-BOK TOOLKIT v1.2 · NL

Volwassenheidskalibratiekit

Aanvulling op de AI Body of Knowledge v1.2

Jan Willem van Veen · ArchiXL · ai-bok.nl

Volwassenheidskalibratiekit: numerieke drempelwaarden + uitgewerkte assessments

Voegt numerieke drempelwaarden en drie uitgewerkte assessments van fictieve organisaties toe om volwassenheidsscore te verankeren en inter-assessor-variantie te verminderen.

1. Wat dit addendum toevoegt

De volwassenheidsbewijs-rubric definieert per KG × niveau de **soorten** bewijs die vereist zijn (Documents, Metrics, Logs). Dit addendum voegt twee ontbrekende onderdelen toe:

- Numerieke drempelwaarden** voor de metric-dragende vereisten - waar de literatuur, regelgeving of operationele ervaring een verdedigbaar getal ondersteunt, wordt de drempelwaarde benoemd. Elke drempelwaarde is getagd als **indicative** (een startpunt, de organisatie mag bijstellen) of **mandatory** (niet onderhandelbaar; niet halen betekent dat het niveau niet is behaald).
- Een kalibratiekit met drie uitgewerkte assessments** van fictieve organisaties op drie verschillende volwassenheidsprofielen, zodat assessoren hun eigen oordeel kunnen ijkken aan gedeelde voorbeelden.

De rubric blijft de bron van waarheid voor *welke soorten* bewijs vereist zijn; dit addendum specificeert *hoeveel*.

2. Numerieke drempelwaarden - wat telt als L2, L3, L4

KG1 - AI Governance

Vereiste	L2	L3	L4	Tag
Goedkeuring AI Governance Charter	in concept	goedgekeurd door AI Board	goedgekeurd door bestuur <i>en</i> jaarlijks herzien	mandatory
AI-inventaris-dekking (% geregistreerde productie-AI-systemen)	≥ 60%	≥ 95%	100% met maandelijks attestatie	mandatory
Cadans jaarlijkse self-assessment	afwezig	jaarlijks	halfjaarlijks	indicative

KG2 - AI Strategie & Portfolio

Vereiste	L2	L3	L4	Tag
Cadans portfolioreview	informeel	per kwartaal	per kwartaal + portfolio-trigger-gebaseerd	mandatory
Meting van waarderealisatie	geen	per systeem ex-post	waarderealisatie-dashboard op portfolioniveau	indicative
Toepassing uitfaseringsprotocol (% uitfaseringen conform protocol)	n.v.t.	≥ 80%	100%	indicative

KG3 - AI Architectuur

Vereiste	L2	L3	L4	Tag
Versie referentiearchitectuur	v1	v1+ met kwartaalreview	continu doorontwikkeld	mandatory
Architectuurconformiteitsassessments (% nieuwe systemen met expliciet conformiteitsrecord)	n.v.t.	≥ 80%	100% met geautomatiseerde checks	indicative
Cognitievlak-elementen (mandaat, bevoegdhedenregister, policy-as-code) operationeel	conceptueel	per agentisch systeem	portfolio-breed	mandatory (indien agentische systemen aanwezig)

KG4 - AI Lifecycle Management

Vereiste	L2	L3	L4	Tag
Lifecycle-gates (minimaal Gates 2, 5, 7)	gedocumenteerd	toegepast op ≥ 90% van wijzigingen	toegepast op 100% met geautomatiseerde gate-logs	mandatory
Gemiddelde tijd gate-besluit → start volgende fase	informeel	≤ 10 werkdagen	≤ 5 werkdagen	indicative
Retrainingsbeleid per model	afwezig	per model gedocumenteerd	geautomatiseerde trigger-pipelines	indicative

KG5 - Data & Semantiek voor AI

Vereiste	L2	L3	L4	Tag
Dekking databron-allowlist	prioriteitsbronnen	alle productiebronnen	alle bronnen incl. dev/test	mandatory
Datakwaliteits-KPI's per bron (compleetheid, actualiteit, nauwkeurigheid)	prioriteitsbronnen	alle productiebronnen	continue monitoring	indicative
Conceptuele raamwerken voor prioriteitsdomeinen	één domein	kerndomein volledig	organisatiebreed	indicative

KG6 - Modelmanagement

Vereiste	L2	L3	L4	Tag
Dekking modelregister	productiemodellen	alle productie + pilot	alle + dev-modellen	mandatory
Modelkaarten (% in-scope systemen met actuele kaart)	≥ 70%	≥ 90%	100%	mandatory
Gemiddelde staleness modelkaart	n.v.t.	≤ 60 dagen	≤ 30 dagen	indicative
Geautomatiseerde bias-detectie in CI/CD	afwezig	voor nieuwe modellen	voor alle modelwijzigingen	indicative

KG7 - AI Interactie & UX

Vereiste	L2	L3	L4	Tag
WCAG-conformiteit voor AI-gerichte UI's	AA voor nieuw	AA voor alle productie	AA + AAA als ambitie	mandatory (waar gereguleerd)
Prompt-pattern-library	basaal	versie-beheerd + getest	A/B-getest per patroon	indicative
Human-in-the-loop-protocollen per risiconiveau	gedocumenteerd	toegepast + gemeten	adaptief	mandatory

KG8 - AI Operations

Vereiste	L2	L3	L4	Tag	
Monitoring-dekking (hoogvolume-systemen met latency-, fout- en drift-metrics)		≥ 50%	≥ 90%	100% + traces	mandatory

Vereiste	L2	L3	L4	Tag
Gemiddelde detectietijd (MTTD) productie-incident	uren	≤ 30 min	≤ 5 min	indicative
Gemiddelde hersteltijd (MTTR)	dagen	≤ 4 uur	≤ 1 uur	indicative
Deployment-automatisering (% deploys via CI/CD)	handmatig	≥ 80%	100% + canary	indicative

KG9 - AI Risicomanagement & Veiligheid

Vereiste	L2	L3	L4	Tag
EU AI Act-risicoclassificatie	voor hoog-risico	voor alle systemen	voor alle + wijziging-getriggerd	mandatory
Cadans actualisatie risicoregister	jaarlijks	per kwartaal	continu	mandatory
Red-team-cadans (hoog-risicosystemen)	n.v.t.	jaarlijks	halfjaarlijks	indicative
FM-coverage-assessment (% van FM01-22 met benoemde control)	≥ 50%	≥ 80%	100%	mandatory (voor agentische landschappen)

KG10 - AI Compliance & Audit

Vereiste	L2	L3	L4	Tag
Algoritmeregister-dekking (NL publieke sector)	hoog-risico	alle in-scope systemen	alle + archief van uitgefaseerde systemen	mandatory (NL publiek)
DPIA-dekking (% hoog-risicosystemen met actuele DPIA)	≥ 80%	100%	100% + wijziging-getriggerd	mandatory
Audit-trail-volledigheid (% besluiten met verifieerbaar spoor)	kernflows	≥ 95%	100%	mandatory
Doorlooptijd conformiteitsassessment	n.v.t.	≤ 90 dagen	≤ 30 dagen	indicative

KG11 - AI Ethiek & Verantwoorde AI

Vereiste	L2	L3	L4	Tag
Bias-audit-dekking (% geauditeerde besluit-beïnvloedende systemen)	≥ 50%	≥ 90%	100% + wijziging-getriggerd	indicative
Cadans ethiekcommissie	jaarlijks	per kwartaal	naar behoefte + per kwartaal	mandatory
Stakeholderparticipatiecycli per jaar	informeel	jaarlijks	continu	indicative

KG12 - Knowledge & Context Management

Vereiste	L2	L3	L4	Tag
Dekking trusted-source-register	prioriteitsbronnen	alle RAG-bronnen	alle kennisinteracties	mandatory
Grounding rate (RAG-systemen: % responses met geciteerde bron)	≥ 70%	≥ 90%	≥ 95%	indicative
Cadans bronactualiteitsaudit	jaarlijks	per kwartaal	continu	indicative

KG13 - AI-Geletterdheid & Capability Development

Vereiste	L2	L3	L4	Tag
Competentieprofielen (% gedefinieerde rol × systeem-doorsnijdingen)	prioriteitsrollen	≥ 90%	100%	mandatory
Actualiteit geletterdheidsregister (% personen actueel op vereiste modules)	≥ 70%	≥ 90%	≥ 95%	mandatory (Art. 4)
Refresh-adherence (% modules binnen refresh-venster)	≥ 70%	≥ 90%	≥ 95%	mandatory
Drill-cadans (supervisors van hoog-risicosystemen)	n.v.t.	jaarlijks	halfjaarlijks	indicative
Gereedheid Artikel 4-bewijsdossier	gestart	onderhouden + auditeerbaar	continu geverifieerd	mandatory

3. Drempelwaarden bijstellen

Indicatieve drempelwaarden mogen door adopterende organisaties met documentatie worden bijgesteld. Verplichte drempelwaarden mogen niet worden versoepeld zonder de niveauclaim ongeldig

te maken. Er bestaat een gedocumenteerd uitzonderingspad (vgl. de aggregatieregels van de rubric), met de volgende structuur:

```
exception_id: EX-2026-Q2-01
ka: KA10
requirement: Audit-trail-volledigheid ≥ 95% op L3
realised: 88%
reason: legacy-classifier zonder backfill; geplande uitfasering Q3 2026
mitigation: out-of-band trace via zaakstelsel; auditor accepteert compenserende maatregel
approved_by: AI Governance Lead + AI Auditor
approved_on: 2026-04-15
expires: 2026-09-30
```

Uitzonderingen worden herzien op de cadans van het inter-assessor-protocol van de rubric. Een uitzondering die langer dan twee reviewcycli voortduurt, verliest haar status (niet langer een uitzondering; wordt óf een afgesloten punt óf een afwaardering).

4. Kalibratiekit - drie uitgewerkte assessments

Drie fictieve organisaties, drie verschillende volwassenheidsprofielen. Een assessor die deze leest, kan de eigen scoring ijkken aan de uitgewerkte voorbeelden.

4.1 Assessment A - "Gemeente Veenwoorden" (≈120k inwoners, publieke sector)

Profiel: net voorbij MVG, op weg naar L3 op de zeven MVG-items.

KG	Score	Bewijssamenvatting
KG1	L3	Charter goedgekeurd maart 2026; jaarlijkse review ingepland; inventaris op 96%
KG2	L2	Portfolioregister aanwezig; kwartaalreviews informeel; nog geen waarderealiserings-dashboards
KG3	L2	Referentiearchitectuur v1.1; cognitievlaak gepositioneerd voor agentische pilot; conformiteitschecks handmatig
KG4	L2	Gates gedocumenteerd; toegepast op 85% van wijzigingen (net onder de L3-drempel van 90%)
KG5	L2	Databronbeleid; één SKOS-raamwerk; datakwaliteits-KPI's voor topbronnen
KG6	L2	Modelregister; modelkaarten op 78% van productie (onder L3 90%)
KG7	L3	WCAG AA over de gehele productie; prompt-pattern-library v0.3 versie-beheerd
KG8	L2	Monitoring op 4/7 hoogvolume-systemen; MTTD ~1 uur; handmatige deploys
KG9	L3	Risicoclassificatie alle systemen; kwartaalregister; één red-team uitgevoerd; FM-coverage 75% (net onder L3 80%) → geflagd , geaccepteerd als L2

KG	Score	Bewijssamenvatting
KG10	L3	Algoritmeregister 100%; DPIA's actueel; audit-trail >95% op vlaggenshipsysteem
KG11	L2	Kader; bias-audits op 3/3 classifiers; ethiekcommissie per kwartaal
KG12	L2	Trusted-source-register; RAG-grounding rate 82% (onder L3 90%)
KG13	L2	Profielen voor 5 rolgroepen; registeractualiteit 78%; refresh-adherence 81%

Aggregaatvolwassenheid: Minimum over alle KG's = **L2**, met KG1, KG7, KG10 op L3. De dominante beperking is het FM-coverage-gat in KG9 (75% vs. 80%); het dichtend daarvan tilt een L3-waardig KG omhoog zonder de totaalscore te beïnvloeden. **Het aggregaat blijft L2** totdat de laagste KG's stijgen.

4.2 Assessment B - "FinanCo NV" (middelgrote financiële dienstverlener, ~5.000 FTE)

Profiel: volwassen op risico en compliance; achter op agentische-AI-integratie.

KG	Score	Toelichting
KG1	L4	Jaarlijkse self-assessment; KPI's in bestuursdashboard; policy-as-code in pilot
KG2	L3	Portfolioreview per kwartaal; waarde per systeem gemeten
KG3	L3	Referentiearchitectuur per kwartaal herzien; cognitievlak gepositioneerd maar nog voor geen enkel systeem operationeel
KG4	L3	Gates toegepast op 94%; CI/CD geautomatiseerd voor ML
KG5	L3	Datakwaliteits-KPI's continu; conceptuele raamwerken voor kerndomein
KG6	L3	Modelkaarten 92%; geautomatiseerde bias-detectie in CI/CD
KG7	L3	A/B-testing + HITL-protocollen per risiconiveau
KG8	L4	Volledige observability; MTTD <5 min; deploys 100% geautomatiseerd
KG9	L4	Red-team per kwartaal; supply-chain-risico's beoordeeld; FM-coverage 100% maar nog geen agentisch systeem
KG10	L4	ISO 42001-gecertificeerd; geautomatiseerde compliance-monitoring
KG11	L3	Bias-audits in de lifecycle; fairness monitoring in opkomst
KG12	L3	Trusted-source-register; grounding rate 91%
KG13	L2	Competentieprofielen voor bestuur + sleutelrollen; registeractualiteit 76%; refresh-adherence 78%; laagste KG

Aggregaatvolwassenheid: L2 (begrensd door KG13). Zelfs met meerdere KG's op L4 houdt de aggregatieregel (minimum over alle KG's) de hoofdscore op L2 totdat KG13 stijgt. **Dit is het juiste**

antwoord volgens de rubric: geletterdheid is structureel kritisch en kan niet worden verhuld achter kracht elders. Een gedocumenteerde uitzondering is mogelijk (gefaseerde uitrol van de Art. 4-gereedheid), maar mag niet permanent worden.

4.3 Assessment C - "AgenticStartUp" (startup van 50 FTE, agentische AI als kernproduct)

Profiel: diepgaand op architectuur en risico voor het agentische vlaggenschip; dun op governance en geletterdheid.

KG	Score	Toelichting
KG1	L2	Charter in concept; AI Board = medeoprichters
KG2	L2	Portfolio = 1 product + 3 interne tools
KG3	L4	Cognitievlak operationeel; bevoegdhedenregister live; ondertekende besluitartefacten; schone variant van de modelleerconventies
KG4	L3	Gates toegepast; CI/CD geautomatiseerd
KG5	L2	Provenance op alle RAG-bronnen; geen conceptueel raamwerk
KG6	L3	Modelkaarten 100%; foundation-model via doorverwijzing gedekt
KG7	L3	UX-patronen versie-beheerd; HITL voor middelrisico
KG8	L3	Observability en CI/CD; kostenmonitoring actief
KG9	L4	FM-coverage 100%; controls per FM operationeel; red-team maandelijks
KG10	L2	Algoritmeregister n.v.t. (geen Nederlandse overheid); EU AI Act-conformiteit in uitvoering
KG11	L2	Ethisch kader v0.5; ethiekcommissie = medeoprichter + adviseur
KG12	L3	Trusted-source-register; grounding rate 93%
KG13	L1	Geen competentieprofielen; geen geletterdheidsregister; ad-hoc-training; verplichte drempelwaarde niet gehaald

Aggregaatvolwassenheid: L1 (KG13 onder de verplichte L2-drempel). Ook al staan KG3 en KG9 op L4, de aggregatieregule van het raamwerk beschouwt dit als L1. **Dit is het juiste oordeel**, ook al voelt het "onjuist" - een organisatie zonder enige geletterdheidsstructuur is *niet* verdedigbaar volwassen, hoe sterk de architectuur ook is. Het resultaat is bedoeld om ongemakkelijk te zijn: til KG13 naar L2 (competentieprofielen voor de 4 sleutelrollen, register gestart, 70% actualiteit) en het aggregaat springt naar L2. Het groeipad is kort en heeft grote hefboomwerking.

5. Hoe de kalibratiekit te gebruiken

1. **Lees eerst alle drie de voorbeelden** voordat u de eigen organisatie scoort.

2. **Kies het voorbeeld dat het dichtst bij het eigen portfolio ligt** qua omvang en sector.
3. **Vergelijk per KG de bewijsvorm en de drempelwaarden.** Gebruik de tabel in §2 om te controleren of de eigen metrics de drempelwaarde halen.
4. **Flag elke grensscore expliciet.** Grensgeval = binnen 10% van een verplichte drempelwaarde of één van de bewijssoorten (Document / Metric / Log) ontbreekt.
5. **Voer inter-assessor-kalibratie uit** boven L3 (rubric): twee assessoren onafhankelijk, peer-review bij oneens.
6. **Pas de aggregatieregel rigoureus toe.** Weersta de verleiding om te middelen - de regel is het *minimum* over alle KG's, tenzij een gedocumenteerde uitzondering het gat dekt.
7. **Documenteer uitzonderingen** zoals in §3; herzie op de cadans van de rubric.

6. Inter-assessor-variantie - bekende valkuilen

- **Overscoring van KG1.** Goedgekeurd charter ≠ werkende governance. L3 vereist bewijs dat de governance *functioneert*.
- **Underscoring van KG8.** Operations kan volwassener zijn dan het lijkt, omdat telemetrie onzichtbaar draait. Vraag naar de metrics in plaats van het runbook.
- **Overscoring van KG10.** Een Algoritmeregister-entry zonder actuele DPIA is L2, niet L3.
- **Overscoring van KG11.** Ethiekcommissie bestaat ≠ ethiekcommissie functioneert. L3 vereist een cyclus van besluiten.
- **Overscoring van KG13.** Artikel 4-bewijsdossier bestaat ≠ registeractualiteit op drempelwaarde. Controleer de actualiteitsmetric.
- **Underscoring van KG3 in niet-agentische landschappen.** De verplichte cognitievlak-items gelden alleen wanneer agentische systemen aanwezig zijn. Pre-agentisch wordt KG3 gescoord tegen de klassieke architectuurcriteria.

7. Relatie met de rubric

Deze kit is *additief* op de rubric. De rubric definieert nog steeds welke soorten bewijs vereist zijn; de kit voegt de drempelwaarden en de uitgewerkte ankers toe. Een organisatie die de voorkeur geeft aan de rubric zonder numerieke drempelwaarden (bijv. omdat het portfolio te klein is voor betekenisvolle percentages) mag alleen de rubric gebruiken - maar neemt dan de variantie op zich die de panelreviewers signaleerden (P8).