Aantonen, niet voldoen

Martijn van Grieken
september 2026

CONTENTS

Waarom de AI-verordening een bewijsregime is, en hoe u dat bewijs uit uw systemen haalt. Een raamwerk voor organisaties die AI inzetten in besluiten over burgers, risico's en patiënten.

Managementsamenvatting

Op 27 juli 2026 trad de Digitale omnibus inzake AI in werking, zes dagen voor de datum waarop de zwaarste verplichtingen van de AI-verordening zouden ingaan. De eisen voor zelfstandige AI-systemen met een hoog risico gelden nu vanaf 2 december 2027; voor AI in gereguleerde producten vanaf 2 augustus 2028. Veel organisaties hebben dat gelezen als een adempauze. Dat is het maar half. Wat is uitgesteld, is de datum waarop een toezichthouder u op de hoog-risico-eisen mag aanspreken; de verboden praktijken, de transparantieplicht en de regels voor AI-modellen voor algemene doeleinden gelden al. Wat niet is uitgesteld, is het vermogen om die vraag te beantwoorden. En dat vermogen laat zich niet achteraf organiseren: een besluit uit 2026 is in 2027 alleen te reconstrueren als het in 2026 is vastgelegd.

De kern van deze whitepaper is een verschuiving van perspectief. De AI-verordening, en de kaders die er al lagen (de AVG, de Algemene wet bestuursrecht, de medische-hulpmiddelenverordening, de rechtspraak over algoritmische besluitvorming), vragen niet in de eerste plaats dat u voldoet. Ze vragen dat u kunt láten zien dat u voldoet, aan een forum dat vragen stelt die u nu nog niet kent, over een moment dat voorbij is. Voldoen is een toestand. Aantonen is een vermogen. En dat vermogen is geen juridische eigenschap van uw organisatie, maar een technische eigenschap van uw systemen: het zit in versiebeheer, logging, herkomst van data en de vastlegging van menselijke tussenkomst.

Wij vertalen dat naar een raamwerk. Elk forum stelt uiteindelijk drie vragen: wat was het systeem, wat deed het, en wie besliste? En het stelt die vragen op twee momenten: vooraf, over het systeem zoals het bedoeld was, en achteraf, over het systeem zoals het wás. Dat levert een bewijsmatrix van zes cellen op. De meeste compliance-programma's vullen de linkerkolom: documenten, beoordelingen, verklaringen. Toezichthouders, rechters en bezwaarcommissies stellen hun vragen in de rechterkolom. Twee mechanismen maken die kloof structureel: bewijsverval, doordat AI-systemen veranderen zonder dat iemand daartoe besluit, en de toezichtsillusie, doordat menselijke controle op papier bestaat maar in de praktijk meebeweegt met het systeem. Beide zijn meetbaar. Wij herleiden ze tot vijf

auditgrootheden en sluiten af met vijf dingen die u morgen kunt doen, te beginnen met één toets: kies een besluit van vorig kwartaal en probeer het te reconstrueren. Hoe lang dat duurt, is het eerlijkste getal dat u over uw AI-governance kunt krijgen.

Juist in publieke uitvoering, veiligheid en gereguleerde zorg is dit geen compliance-kwestie in de marge. Daar is het antwoord op “hoe kwam dit besluit tot stand?” een recht van de burger, de verdachte of de patiënt, en de plicht om dat antwoord te kunnen geven bestond al vóór de AI-verordening.

---

1. De deadline is verschoven, de bewijslast niet

Eerst de feiten, omdat ze in 2026 zijn verschoven. De AI-verordening (Verordening (EU) 2024/1689) trad op 1 augustus 2024 in werking en wordt gefaseerd van toepassing. De verplichtingen voor AI-systemen met een hoog risico zouden op 2 augustus 2026 ingaan. Omdat de geharmoniseerde normen en de nationale toezichtstructuren daar niet klaar voor waren, stelde de Europese Commissie in november 2025 de Digitale omnibus inzake AI voor. Na een politiek akkoord in mei en instemming van Parlement en Raad in juni is die op 24 juli 2026 gepubliceerd als Verordening (EU) 2026/1744, en op 27 juli in werking getreden.

Het tijdpad, per 1 september 2026:

Datum	Wat geldt
Sinds 2 februari 2025	Verboden AI-praktijken (art. 5). AI-geletterdheid (art. 4); de omnibus verving die verplichting op 27 juli 2026 door een inspanningsverplichting
Sinds 2 augustus 2025	Verplichtingen voor AI-modellen voor algemene doeleinden; governance en sancties
Sinds 2 augustus 2026	Transparantieverplichtingen (art. 50): wie met een AI-systeem praat moet dat weten, en synthetische content, deepfakes en emotieherkenning moeten kenbaar zijn. Voor het markeren van synthetische content geldt tot 2 december 2026 een overgangstermijn voor systemen die al op de markt waren
2 december 2026	Nieuwe verboden op AI die zonder toestemming intiem beeldmateriaal of materiaal van kindermisbruik genereert
2 december 2027	Verplichtingen voor zelfstandige AI-systemen met een hoog risico (bijlage III); was 2 augustus 2026
2 augustus 2028	Verplichtingen voor hoog-risico-AI in gereguleerde producten (bijlage I), waaronder medische hulpmiddelen; was 2 augustus 2027
2 augustus 2030	Uiterste datum voor hoog-risicosystemen die al draaiden: zie de overgangsregel hieronder (art. 111)

Let op de overgangsregel voor wat er al draait. Een hoog-risicosysteem dat vóór de toepassingsdatum op de markt was of in gebruik was genomen, valt pas onder de eisen zodra het wezenlijk wordt gewijzigd (art. 111). Die uitzondering geldt niet onbeperkt: aanbieders en gebruiksverantwoordelijken van hoog-risicosystemen die bedoeld zijn voor gebruik door overheidsinstanties moeten uiterlijk 2 augustus 2030 voldoen. Voor publieke uitvoering is dat dus geen vrijstelling maar een langere aanloop, en precies daarin moet het bewijs ontstaan.

In Nederland bracht het kabinet in april 2026 de Uitvoeringswet AI-verordening in consultatie. Die belegt het toezicht bij tien bestaande toezichthouders, elk in het eigen domein, met de Autoriteit Persoonsgegevens en de Rijksinspectie Digitale Infrastructuur als coördinatoren. Dat de nationale wet later komt dan de Europese verplichtingen, verandert niets aan die verplichtingen: een verordening werkt rechtstreeks. En wie de toezichthouder is, verandert niets aan wat u moet kunnen laten zien.

Wie in het uitstel uitsluitend zestien maanden extra leest, mist drie dingen.

Ten eerste: de bewijsplicht bestond al. Artikel 22 van de AVG geeft betrokkenen het recht niet te worden onderworpen aan een uitsluitend geautomatiseerd besluit met rechtsgevolg of vergelijkbaar aanmerkelijk effect, en verlangt waar zo een besluit wél is toegestaan passende waarborgen, waaronder het recht op menselijke tussenkomst; artikel 35 verlangt een gegevensbeschermingseffectbeoordeling. De Algemene wet bestuursrecht eist dat een besluit berust op een deugdelijke en kenbare motivering. En de Nederlandse rechter heeft dat al tweemaal expliciet op algoritmen toegepast. In 2017 oordeelde de Afdeling bestuursrechtspraak van de Raad van State over het rekenmodel AERIUS dat een bestuursorgaan dat een besluit op een computermodel baseert, de gemaakte keuzes, gebruikte gegevens en aannames volledig, tijdig en uit eigen beweging openbaar moet maken, omdat het besluit anders niet controleerbaar en aanvechtbaar is. In 2020 verklaarde de rechtbank Den Haag de wettelijke regeling van het fraudesignaleringssysteem SyRI onverbindend wegens strijd met artikel 8 EVRM, mede omdat de werking ervan onvoldoende transparant en verifieerbaar was (Wieringa, 2023). Beide uitspraken gingen niet over de vraag óf de overheid iets fout deed, maar over de vraag of dat te controleren viel.

Ten tweede: bewijs is alleen vooruit op te bouwen. De AI-verordening verplicht gebruiksverantwoordelijken de automatisch gegenereerde logs van een hoog-risicosysteem ten minste zes maanden te bewaren (art. 26, lid 6), en aanbieders om technische documentatie actueel te houden (art. 11). Maar de eigenlijke reden om nu te beginnen is niet die bewaartermijn. Het is dat elk besluit dat vandaag zonder vastgelegde data- en modelversie wordt genomen, in december 2027 definitief onreconstrueerbaar is. Uitstel van de handhavingsdatum is geen uitstel van het moment waarop uw bewijs begint te ontstaan.

Ten derde: de achterstand is empirisch vastgesteld. In 2024 inventariseerde de Algemene Rekenkamer 433 AI-systemen bij 70 rijksorganisaties. Voor meer dan de helft was geen risicoafweging gemaakt; bij 35 procent van de systemen in gebruik wist de organisatie niet of het systeem de verwachting waarmaakte; 5 procent stond in het openbare Algoritmeregister. Twee jaar eerder had de Rekenkamer bij zes van negen onderzochte algoritmen risico’s gevonden, van gebrekkige controle op prestaties tot vooringenomenheid en ongeautoriseerde toegang (Algemene Rekenkamer, 2022, 2024). Dat is geen beeld van organisaties die de regels overtreden. Het is een beeld van organisaties die niet kunnen laten zien wat ze doen.

2. Verantwoording is een relatie, geen document

Het woord “compliance” stuurt het denken de verkeerde kant op. Het suggereert een toestand waarin een organisatie zich bevindt of niet, vast te stellen met een checklist. Wat de wet werkelijk vraagt, is beter beschreven in de bestuurskunde. Verantwoording, in de definitie van Mark Bovens, is een relatie tussen een actor en een forum, waarin de actor verplicht is zijn handelen uit te leggen en te rechtvaardigen, het forum vragen kan stellen en een oordeel kan vellen, en de actor consequenties kan ondervinden (Bovens, 2007). Drie elementen maken die definitie scherp voor AI. Het forum bepaalt de vragen, niet de actor. De vragen komen achteraf. En het oordeel volgt niet op wat de actor bedoelde, maar op wat hij kan laten zien.

Voor algoritmische systemen komt daar een complicatie bij. Uit een systematische studie van 242 publicaties over algoritmische verantwoording concludeert Wieringa (2020) dat de “actor” bij een algoritme zelden één persoon is, maar een netwerk van ontwikkelaars, leveranciers, beheerders en gebruikers, verdeeld over tijd. Wie op het moment van de vraag aan tafel zit, heeft het systeem meestal niet gebouwd en vaak niet zien veranderen. Verantwoording afleggen betekent dan: namens dat hele netwerk kunnen reconstrueren wat er is gebeurd.

Dat is ook waarom transparantie het probleem niet oplost. Ananny en Crawford (2018) laten zien dat inzicht in een systeem, tot en met de broncode, iets anders is dan het vermogen om het te bevragen; wie een model kan zien, weet nog niet wat het op een gegeven dag met een gegeven dossier heeft gedaan. Kroll en collega’s (2017) trekken de consequentie: verantwoording moet in het ontwerp van een systeem worden ingebouwd, met vastlegging waarmee achteraf kan worden bewezen dat een besluit volgens de aangekondigde procedure tot stand kwam. De AI-verordening zegt in wezen hetzelfde: een hoog-risicosysteem moet technisch zo zijn ontworpen dat gebeurtenissen gedurende de levensduur automatisch worden geregistreerd, met het oog op traceerbaarheid (art. 12).

In ons vraagstuk “We moeten voldoen aan nieuwe eisen” formuleerden we vier vragen die een organisatie zichzelf kan stellen: weet u welke data waar staat, kunt u een besluit reconstrueren, is er iemand verantwoordelijk, en waar zit menselijke tussenkomst. Wie de vragen van toezichthouders, rechters, accountants en bezwaarcommissies over algoritmische besluiten naast elkaar legt, ziet dat ze zich laten herleiden tot drie. **Wat was het?** Welk systeem, welke versie, welke data, welke configuratie. **Wat deed het?** Welke uitkomst, op welke invoer, en gedroeg het zich zoals het was gevalideerd. **Wie besliste?** Wie was verantwoordelijk, wie keek

mee, met welk mandaat, en week die persoon af. En elk forum stelt die drie vragen op twee momenten: vooraf, over het systeem zoals het bedoeld was, en achteraf, over het systeem zoals het was.

3. De bewijsmatrix: drie vragen, twee momenten

Kruist u de drie vragen met de twee momenten, dan ontstaan zes soorten bewijs. Tussen haakjes de bepalingen van de AI-verordening waar dat bewijs op terugvalt; de opsomming is illustratief, niet uitputtend; enkele cellen vragen meer vastlegging dan de bepaling letterlijk eist, en welke bepalingen op uw systeem van toepassing zijn, bepaalt uw jurist.

	Vooraf: het systeem zoals bedoeld	**Achteraf: het systeem zoals het was**
Wat was het?	Technische documentatie, datagovernance, beschrijving van trainings- en testdata, registratie in de EU-databank (art. 10, 11, 49)	Welke versie van model, data en configuratie draaide op die dag, en is die toestand terug te halen (art. 12, 18)
Wat deed het?	Validatieresultaten, nauwkeurigheid, robuustheid, bekende beperkingen en voorzienbaar misbruik (art. 9, 15)	Logs van invoer en uitkomst, monitoring op drift, gemelde incidenten (art. 12, 19, 26 lid 6, 72, 73)
Wie besliste?	Rolverdeling aanbieder en gebruiksverantwoordelijke, toezicht toegewezen aan personen met competentie en mandaat, grondrechteneffectbeoordeling (art. 14, 26 lid 2, 27)	Per besluit: wie keek mee, wat zag die persoon, week hij af en waarom; uitleg aan de betrokkene (art. 26 lid 11, 86; art. 22 AVG)

De linkerkolom is waar de meeste compliance-programma's hun energie in steken. Dat is begrijpelijk: het zijn documenten, ze zijn af te vinken, ze passen bij een project met een einddatum, en de conformiteitsbeoordeling vraagt er expliciet om. Het is ook de kolom waar de markt van registers, assessments en certificeringen zich op richt. Maar een forum dat een individueel besluit onderzoekt, stelt vrijwel altijd een vraag uit de rechterkolom. Een bezwaarcommissie wil niet weten welke nauwkeurigheid het model bij oplevering had, maar welke uitkomst het gaf voor deze aanvrager, op

basis van welke gegevens, en of de behandelaar daar nog naar heeft gekeken. Een dossier beschrijft het systeem zoals het bedoeld was; een toezichthouder vraagt naar het systeem zoals het was, op die dag, voor die persoon, met die versie.

De praktische les van de matrix is even simpel als ongemakkelijk: de linkerkolom kunt u met terugwerkende kracht op orde brengen, de rechterkolom niet. Ontbrekende documentatie kan worden geschreven. Ontbrekende logs kunnen niet worden gegenereerd. Wie de rechterkolom pas invult wanneer de handhaving begint, heeft tegen die tijd een dossier vol beloften en geen enkel bewijs.

Daaruit volgt de toets die wij bij elke organisatie als eerste doen, en die u ook zelf kunt doen. Kies een besluit van een willekeurige datum in het vorige kwartaal waarbij een AI-systeem een rol speelde. Probeer, met de mensen en systemen die u nu hebt, boven water te krijgen welke data erin ging, welke versie van het model draaide, welke uitkomst eruit kwam, en wie er vervolgens wat mee deed. Meet hoe lang dat duurt. Wij noemen dit de reconstructietoets. Het resultaat is meestal niet “onmogelijk”, maar “drie weken en vier mensen”, en dat getal is uw uitgangspositie.

4. Twee mechanismen die bestuurders moeten kennen

Bewijsverval: het systeem verandert zonder dat iemand daartoe besluit. Bij een klassieke applicatie is een gedragsverandering vrijwel altijd terug te voeren op een ingreep: een release, een configuratiewijziging, een bijgewerkte afhankelijkheid. Bij een AI-systeem niet. De data die het systeem te zien krijgt, verschuift: populaties, formulieren, registratiepraktijken en beleid veranderen, en daarmee de verdeling waarop het model ooit werd gevalideerd. De literatuur onderscheidt daarbij de verschuiving van de invoerverdeling (dataset shift) van de verschuiving in het verband tussen invoer en uitkomst (concept drift); in de praktijk lopen ze door elkaar en is het gevolg hetzelfde (Gama, Žliobaitė, Bifet, Pechenizkiy, & Bouchachia, 2014) en is in de kliniek beschreven als reden waarom een goedgekeurd model na verloop van tijd stil verslechtert (Finlayson et al., 2021). Modellen worden hertraind, soms automatisch. Bovenstroomse bronnen wijzigen van definitie zonder dat de afnemer het merkt; Sculley en collega’s (2015) noemden dit de verborgen technische schuld van machine learning, samengevat als: wie iets verandert, verandert alles. En wie op een extern gehost taalmodel bouwt, krijgt nieuwe modelversies van de leverancier, zonder eigen release en soms zonder aankondiging. De gevolgtrekking is streng: de afstand tussen het gedocumenteerde en het draaiende systeem groeit vanzelf en

wordt alleen kleiner door actief ingrijpen. Documentatie is geen voorraad die u eenmalig aanlegt, maar een stroom die u moet bijhouden. De verordening is daar expliciet over: risicobeheer is een continu, iteratief proces over de hele levensduur (art. 9), documentatie moet actueel blijven (art. 11), en aanbieders moeten monitoring na het in de handel brengen inrichten (art. 72). Een eenmalige beoordeling bij livegang voldoet aan de letter en mist de geest.

De toezichtsillusie: de mens in de lus beweegt mee. Menselijk toezicht is de veiligheidsklep van vrijwel elk AI-kader: artikel 22 AVG spreekt van menselijke tussenkomst, artikel 14 van de AI-verordening van menselijk toezicht, en bijna elke organisatie heeft een behandelaar die “ernaar kijkt”. Het probleem is dat mensen die dagelijks met een goed presterend systeem werken, de neiging ontwikkelen het te volgen, ook wanneer het fout zit. Dat verschijnsel, automation bias, is sinds eind jaren negentig experimenteel gedocumenteerd (Skitka, Mosier, & Burdick, 1999) en blijkt hardnekkig: het neemt toe met werkdruk en routine, en laat zich met een waarschuwing alleen niet wegnemen (Parasuraman & Manzey, 2010). Green (2022) onderzocht beleid dat menselijk toezicht op overheidsalgoritmen voorschrijft en concludeert dat het vooral werkt als legitimatie: het systeem wordt ingezet omdat er een mens naar kijkt, terwijl die mens niet in staat is het te corrigeren.

De wetgever kent dat risico. Artikel 14 eist dat toezichthoudende personen zich bewust blijven van de neiging om automatisch op het systeem te vertrouwen, uitkomsten juist kunnen interpreteren, kunnen besluiten het systeem niet te gebruiken, en kunnen ingrijpen of stoppen; artikel 26 verlangt dat dit toezicht wordt toegewezen aan personen met de competentie, opleiding en bevoegdheid daarvoor. Onderzoek naar wat toezicht effectief maakt, komt tot dezelfde voorwaarden: de toezichthouder moet werkelijk kunnen ingrijpen, werkelijk kunnen zien wat het systeem doet, en zichzelf onder druk kunnen bijsturen (Sterz et al., 2024); en toezicht werkt alleen als het is ingericht vanuit georganiseerd wantrouwen, met bevoegdheden en tegenkrachten, niet als beleefdheidsvorm (Laux, 2024). Meekijken zonder mandaat, tijd of informatie telt niet. Het is bovendien zelden ergens vastgelegd, en dat is de bestuurlijke consequentie: menselijk toezicht dat niet wordt geregistreerd, bestaat voor een forum niet. Een toezichtfunctie die in tienduizend besluiten nooit van het systeem afwijkt, bewijst niet dat het systeem goed is; het is vooral een reden om na te gaan of er werkelijk wordt toegezien. Wie toezicht serieus neemt, meet het.

5. Van zorg naar meetlat: vijf auditgrootheden

Beide mechanismen laten zich herleiden tot een klein aantal eigenschappen van een uitgerold systeem die u kunt meten, in een audit, maar ook gewoon elk kwartaal in het beheer. Geen filosofie, maar grootheden.

1. **Reconstructietijd.** De tijd die het kost om voor één besluit van een willekeurige datum de invoerdata, de model- en configuratieversie, de uitkomst en de menselijke tussenkomst boven water te krijgen. Streefwaarde: uren, niet weken. Dit is de uitkomst van de reconstructietoets, en de enige grootheid die de hele matrix in één getal vangt.
2. **Spoordekking.** Het aandeel besluiten waarvoor een volledige, onveranderlijke keten bestaat van invoer via versie en uitkomst tot mens. Honderd procent is de norm; elke uitzondering is een besluit dat u niet kunt verantwoorden.
3. **Documentatie-achterstand.** De tijd tussen de laatste wijziging van model, data of pijplijn en de bijgewerkte documentatie. Nul is haalbaar wanneer documentatie uit de pijplijn wordt gegenereerd, in de vorm van modelkaarten en datasheets (Mitchell et al., 2019; Gebru et al., 2021), in plaats van erover te worden geschreven.
4. **Driftdetectietijd.** De tijd tussen het ontstaan van een verschuiving in invoerverdeling of prestaties en het moment waarop iemand het weet. Wie deze grootheid niet kent, ontdekt drift via de klachten.
5. **Afwijkingsgraad.** Hoe vaak toezichthoudende personen van de systeemuitkomst afwijken, en hoe dat is verdeeld over personen, teams en perioden. Nul is een rode vlag. Een zeer hoge waarde ook: dan wordt het systeem niet vertrouwd of is het niet bruikbaar. Interessant is de bandbreedte daartussen, en de vraag of afwijkingen worden gemotiveerd.

Deze vijf grootheden zijn eigenschappen van uw architectuur, niet van uw juridische documentatie. Ze horen thuis in de beheerfase van elke AI-toepassing, naast de vertrouwde monitoring op prestaties en beveiliging. En het zijn precies de getallen die een toezichthouder, een accountant of een rechter uiteindelijk nodig heeft om de drie vragen uit hoofdstuk 2 beantwoord te krijgen.

6. Wat kan een organisatie morgen doen

Vijf handelingsperspectieven volgen rechtstreeks uit het raamwerk. Ze zijn geschreven voor organisaties waar AI-systemen meewerken aan besluiten over burgers, risico’s en patiënten, en waar de vraag “hoe kwam dit tot stand?” dus geen

hypothetische is.

1. **Maak de inventaris compleet en geef elk systeem een naam.** Welke AI-systemen zijn er, in welke rol (aanbieder of gebruiksverantwoordelijke; wie een ingekocht systeem wezenlijk aanpast of onder eigen naam uitbrengt, kan aanbieder worden), en in welke risicoklasse. De classificatie is werk voor uw jurist; de inventaris is werk voor IT en business samen, en die twee moeten in dezelfde lijst kijken. Koppel per systeem één verantwoordelijke met naam. Gedeelde verantwoordelijkheid betekent in de praktijk dat niemand het overzicht heeft op het moment dat het nodig is.
2. **Doe de reconstructietoets, nu.** Eén besluit, één willekeurige datum, één stopwatch. Het resultaat is uw nulmeting op alle vijf de grootheden tegelijk, en in onze ervaring het meest overtuigende argument dat ooit in een stuurgroep over dit onderwerp is ingebracht. Wat nu drie weken kost, moet in december 2027 drie uur kosten.
3. **Bouw herleidbaarheid in de architectuur, niet in het dossier.** Versiebeheer van data, modellen, configuraties en prompts; onveranderlijke logs met een bewaartermijn die ten minste de wettelijke zes maanden dekt en in de publieke sector doorgaans veel langer moet, gelet op bezwaar- en archieftermijnen; herkomst van data die door de hele keten wordt meegedragen; documentatie die uit de pijplijn wordt gegenereerd. Dit is dezelfde laag waarop ook het vertrouwen in uw cijfers rust: definities, herkomst en eigenaarschap. Wie die laag voor verantwoording bouwt, bouwt haar ook voor datakwaliteit, en andersom.
4. **Geef toezicht tanden en meet het.** Leg per hoog-risicotoepassing vast wie toezicht houdt, met welk mandaat om af te wijken, met welke opleiding, en met hoeveel tijd per besluit. Registreer elk oordeel van de toezichthoudende persoon als eigen gebeurtenis, inclusief afwijkingen en hun motivering. Ontwerp frictie waar dat past: interfaces die de behandelaar eerst zelf laten oordelen voordat het systeem zijn suggestie toont, verminderen aantoonbaar het blind overnemen (Buçinca, Malaya, & Gajos, 2021). En bespreek de afwijkingsgraad periodiek, niet als prestatiemaat van medewerkers, maar als gezondheidsmaat van het toezicht.
5. **Behandel bewijs als een stroom, niet als een stuk.** Koppel documentatie aan het releaseproces, zodat een wijziging zonder bijgewerkte documentatie niet kan worden uitgerold. Richt monitoring na livegang in op drift en incidenten, met een meldroute die de meldplicht voor ernstige incidenten (art. 73) kan bedienen. Plan hervalidatie op een vaste cadans en na elke wezenlijke wijziging. En beleg dit alles in beheer, met budget, niet in een project dat eindigt bij livegang. De conformiteitsbeoordeling is het begin van het bewijs, niet het einde.

7. Drie sectoren, drie accenten

Publieke uitvoering. Hier is verantwoording geen nieuwe verplichting maar de kern van het vak: bestuursorganen motiveren besluiten, burgers maken bezwaar, en de rechter toetst. De AI-verordening voegt daar voor overheden een grondrechteneffectbeoordeling vóór ingebruikname van een hoog-risicosysteem aan toe (art. 27), en, voor de meeste toepassingen uit bijlage III, het recht van betrokkenen op uitleg bij een besluit met rechtsgevolg of vergelijkbaar aanmerkelijk effect (art. 86). Nederland liep hierin vooruit met het Algoritmeregister, het Impact Assessment Mensenrechten en Algoritmes (Gerards, Schäfer, Vankan, & Muis, 2021) en het Algoritmekader van het ministerie van BZK; de Rekenkamercijfers laten zien dat de instrumenten er zijn en de vulling achterblijft. Het diepere punt is ouder dan al deze kaders: Bovens en Zouridis (2002) beschreven al hoe discretie in uitvoeringsorganisaties verschuift van de behandelaar naar het systeem, en daarmee de plek waar verantwoording moet worden georganiseerd. Wie die verschuiving niet volgt, krijgt uitvoering waarin niemand meer kan navertellen waarom een burger een besluit ontving.

Veiligheid. In het veiligheidsdomein staat een reeks toepassingen op de lijst met een hoog risico, van risico-inschatting tot de beoordeling van de betrouwbaarheid van bewijs, en is voorspellend politiewerk dat uitsluitend op profilering berust verboden. Tegelijk zijn logs hier gevoeliger dan waar ook, en gelden eigen regimes voor registratie en openbaarheid. Dat maakt de rechterkolom van de matrix niet minder belangrijk, maar juist bepalender: waar de inzet van een systeem niet openbaar kan zijn, is reconstrueerbaarheid achteraf, voor een rechter, een toezichthouder of een intern controleorgaan, het enige dat overblijft. Meijer en Wessels (2019) laten in hun overzicht van voorspellend politiewerk zien dat de beloofde effectiviteit zelden hard is aangetoond, terwijl de risico’s voor grondrechten dat wel zijn. Verantwoording is hier dus ook de manier om te bewijzen dat een systeem werkt.

Gereguleerde zorg. AI die als medisch hulpmiddel kwalificeert, valt al onder de medische-hulpmiddelenverordening en wordt onder de AI-verordening een hoog-risicosysteem uit bijlage I, met de nieuwe datum van 2 augustus 2028 (Van Kolfschooten & Van Oirschot, 2024). Voor fabrikanten betekent dat één technisch dossier voor twee regimes. Voor zorginstellingen als gebruiksverantwoordelijke is de crux dat de klinische validatie een momentopname is, en dat een model in de praktijk aan dataset shift onderhevig is (Finlayson et al., 2021): andere scanners, andere populaties, andere registratie. De zorg heeft met post-market surveillance en

calamiteitenmeldingen al een cultuur van bewijs na livegang. De vraag is of die cultuur ook de modelversies, de invoerdata en het oordeel van de arts omvat, of alleen het apparaat.

8. Waar wij wel en niet over gaan, en een uitnodiging

Twentynext geeft geen juridisch advies en bepaalt niet welke kaders op uw organisatie van toepassing zijn, in welke risicoklasse uw systemen vallen of welke rol u in de keten vervult. Dat hoort bij uw jurist of compliance-functie. Wat wij doen, is de technische kant: zorgen dat de vragen die zij, en straks de toezichthouder, stellen ook uit uw systemen te beantwoorden zijn. De bewijsmatrix en de vijf auditgrootheden zijn daarvoor ons instrument, en de reconstructietoets is waar wij beginnen.

Wilt u weten waar uw organisatie staat? Wij voeren de reconstructietoets uit op één systeem naar keuze en vertalen de uitkomst naar een meetlat op de vijf grootheden, met een volgorde van wat het eerst moet. Van eerste audit tot inrichting en structureel beheer. Neem contact op via info@twentynext.nl.

Over de auteur

Martijn van Grieken is directeur Data & AI bij Twentynext, het Eindhovense data- en AI-consultancybureau dat sinds 2014 werkt voor onder meer publieke uitvoering, veiligheid en gereguleerde zorg. Hij werkt ruim tien jaar op het snijvlak van datagovernance, AI en publieke besluitvorming.

Over Twentynext

Twentynext is een data- en AI-consultancy uit Eindhoven, opgericht in 2014, en ondersteunt organisaties in onder meer publieke uitvoering, veiligheid en gereguleerde zorg, van detachering van specialisten tot projecten met vooraf afgesproken opleverpunten en structureel beheer.

Verantwoording

Deze whitepaper beschrijft de stand van de regelgeving per 1 september 2026. Verwijzingen naar artikelen van de AI-verordening betreffen Verordening (EU) 2024/1689 zoals gewijzigd door Verordening (EU) 2026/1744; de weergave is bedoeld als oriëntatie voor beslissers en is geen juridisch advies. De bewijsmatrix, de vijf auditgrootheden en de reconstructietoets zijn instrumenten uit onze eigen praktijk; de onderliggende begrippen (verantwoording als relatie tussen actor en forum, automation bias, concept drift, verborgen technische schuld) komen uit de aangehaalde literatuur. Alle aangehaalde bronnen zijn door de auteur geverifieerd.

Bij het opstellen van dit stuk is gebruikgemaakt van AI-taalmodellen onder redactie en verantwoordelijkheid van de auteur.

Literatuur

Algemene Rekenkamer. (2022). *Algoritmes getoetst: De inzet van 9 algoritmes bij de rijksoverheid*. Algemene Rekenkamer. https://www.rekenkamer.nl/documenten/2022/05/18/algoritmes-getoetst

Algemene Rekenkamer. (2024). *Focus op AI bij de rijksoverheid*. Algemene Rekenkamer. https://www.rekenkamer.nl/documenten/2024/10/16/focus-op-ai-bij-de-rijksoverheid

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society, 20*(3), 973–989. https://doi.org/10.1177/1461444816676645

Bovens, M. (2007). Analysing and assessing accountability: A conceptual framework. *European Law Journal, 13*(4), 447–468. https://doi.org/10.1111/j.1468-0386.2007.00378.x

Bovens, M., & Zouridis, S. (2002). From street-level to system-level bureaucracies: How information and communication technology is transforming administrative discretion and constitutional control. *Public Administration Review, 62*(2), 174–184. https://doi.org/10.1111/0033-3352.00168

Buçinca, Z., Malaya, M. B., & Gajos, K. Z. (2021). To trust or to think: Cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction, 5*(CSCW1), 1–21.

https://doi.org/10.1145/3449287

Finlayson, S. G., Subbaswamy, A., Singh, K., Bowers, J., Kupke, A., Zittrain, J., Kohane, I. S., & Saria, S. (2021). The clinician and dataset shift in artificial intelligence. *New England Journal of Medicine, 385*(3), 283–286. https://doi.org/10.1056/NEJMc2104626

Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM Computing Surveys, 46*(4), Article 44. https://doi.org/10.1145/2523813

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., & Crawford, K. (2021). Datasheets for datasets. *Communications of the ACM, 64*(12), 86–92. https://doi.org/10.1145/3458723

Gerards, J., Schäfer, M. T., Vankan, A., & Muis, I. (2021). *Impact Assessment Mensenrechten en Algoritmes*. Universiteit Utrecht, in opdracht van het Ministerie van Binnenlandse Zaken en Koninkrijksrelaties.

Green, B. (2022). The flaws of policies requiring human oversight of government algorithms. *Computer Law & Security Review, 45*, 105681. https://doi.org/10.1016/j.clsr.2022.105681

Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2017). Accountable algorithms. *University of Pennsylvania Law Review, 165*(3), 633–705.

Laux, J. (2024). Institutionalised distrust and human oversight of artificial intelligence: Towards a democratic design of AI governance under the European Union AI Act. *AI & Society, 39*(6), 2853–2866. https://doi.org/10.1007/s00146-023-01777-z

Meijer, A., & Wessels, M. (2019). Predictive policing: Review of benefits and drawbacks. *International Journal of Public Administration, 42*(12), 1031–1039. https://doi.org/10.1080/01900692.2019.1575664

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19)*, 220–229. https://doi.org/10.1145/3287560.3287596

Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors, 52*(3), 381–410. https://doi.org/10.1177/0018720810376055

Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J.-F., & Dennison, D. (2015). Hidden technical debt in machine learning systems. *Advances in Neural Information Processing Systems, 28*, 2503–2511.

Skitka, L. J., Mosier, K. L., & Burdick, M. (1999). Does automation bias decision-making? *International Journal of Human-Computer Studies, 51*(5), 991–1006. https://doi.org/10.1006/ijhc.1999.0252

Sterz, S., Baum, K., Biewer, S., Hermanns, H., Lauber-Rönsberg, A., Meinel, P., & Langer, M. (2024). On the quest for effectiveness in human oversight: Interdisciplinary perspectives. *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24)*, 2495–2507. https://doi.org/10.1145/3630106.3659051

Van Kolfschooten, H., & Van Oirschot, J. (2024). The EU Artificial Intelligence Act (2024): Implications for healthcare. *Health Policy, 149*, 105152. https://doi.org/10.1016/j.healthpol.2024.105152

Wieringa, M. (2020). What to account for when accounting for algorithms: A systematic literature review on algorithmic accountability. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20)*, 1–18. https://doi.org/10.1145/3351095.3372833

Wieringa, M. (2023). "Hey SyRI, tell me about algorithmic accountability": Lessons from a landmark case. *Data & Policy, 5*, e2. https://doi.org/10.1017/dap.2022.39

WET- EN REGELGEVING EN RECHTSPRAAK

Verordening (EU) 2024/1689 van het Europees Parlement en de Raad van 13 juni 2024 tot vaststelling van geharmoniseerde regels betreffende artificiële intelligentie (AI-verordening). *PB L*, 2024/1689, 12.7.2024. http://data.europa.eu/eli/reg/2024/1689/oj

Verordening (EU) 2026/1744 van het Europees Parlement en de Raad van 8 juli 2026 tot wijziging van de Verordeningen (EU) 2024/1689, (EU) 2018/1139 en (EU) 2023/1230 wat betreft vereenvoudiging van de uitvoering van geharmoniseerde

regels betreffende artificiële intelligentie (Digitale omnibus inzake AI). *PB L*, 2026/1744, 24.7.2026. http://data.europa.eu/eli/reg/2026/1744/oj

Verordening (EU) 2016/679 (Algemene verordening gegevensbescherming), in het bijzonder de artikelen 22 en 35.

Verordening (EU) 2017/745 betreffende medische hulpmiddelen.

Afdeling bestuursrechtspraak van de Raad van State, 17 mei 2017, ECLI:NL:RVS:2017:1259 (AERIUS).

Rechtbank Den Haag, 5 februari 2020, ECLI:NL:RBDHA:2020:865 (SyRI).

Ministerie van Economische Zaken. (2026). *Uitvoeringswet verordening artificiële intelligentie* (consultatieversie, april 2026). https://www.internetconsultatie.nl/uaiv/b1