Wanneer modellen beleid dragen

Een gespreksinstrument voor legitiem en verantwoord modelgebruik in beleid en uitvoering



Universiteit Leiden

Colofon en inhoudsopgave

Met subsidie van het Ministerie van Binnenlandse Zaken en Koninkrijksrelaties

Opgesteld door: The Hague Centre for Digital Governance, Universiteit Leiden

Auteurs: Friso Selten, David Lake, Sarah Giest, Alex Ingrams & Bram Klievink

Inhoudsopgave:

Hoofdlijnen van het rapport 3
1. Modelmatig besturen 5
2. Een bestuurskundig perspectief op modelgebruik 8
2.1 Verschillende rollen van modellen in beleid 8
2.2 Hoe modeluitkomsten bestuurlijke betekenis krijgen 9
2.3 Modelgebruik is politiek 11
2.4 Van standaard naar legitimiteit, assurance en vertrouwen 13
3. Zes bestuurlijke afwegingen bij modelgebruik 15
3.1 Stap 1: Is modellering een passende manier om het beleidsvraagstuk te ondersteunen? 16
3.2 Stap 2: Op welke kennisbasis zijn de uitkomsten gebaseerd? 17
3.3 Stap 3: Welke betekenis kan aan de uitkomsten worden toegekend? 19
3.4 Stap 4: Hoe zijn verantwoordelijkheid, menselijke tussenkomst en tegenspraak georganiseerd? 20
3.5 Stap 5: Welke informatie is nodig om het gebruik te begrijpen en te beoordelen? 21
3.6 Stap 6: Wanneer beoordelen we opnieuw of het model nog passend is? 22
3.7 De zes afwegingen in samenhang 23
4. Het gespreksinstrument in de praktijk 24
4.1 Wie praten er mee? 24
4.2 Wanneer wordt het gesprek gevoerd? 24
4.3 Institutionele inbedding 25
5. Conclusie 27
Bijlage 1. Gespreksinstrument voor verantwoord modelgebruik 28
Bijlage 2. Totstandkoming, scope en verhouding tot de motie 32
Referenties en geraadpleegde bronnen 35

Hoofdlijnen van het rapport

Modellen zoals rekenmodellen en algoritmen spelen een substantiële rol in beleid, uitvoering en toezicht. Modellen ondersteunen de overheid bij het begrijpen van maatschappelijke vraagstukken, het verkennen van toekomstige ontwikkelingen, het verdelen van schaarse middelen, het prioriteren van toezicht en het onderbouwen van besluiten. Tegelijkertijd zijn modellen steeds vaker onderwerp van politiek en maatschappelijk debat. Niet alleen de technische kwaliteit van modellen staat daarbij ter discussie, maar ook de rol die modeluitkomsten krijgen in beleid en besluitvorming.

Dat is een goede ontwikkeling. Modellen leveren geen objectieve waarheden op waarmee politieke discussies kunnen worden beslecht. Juist doordat zij complexe aspecten van de werkelijkheid selecteren, structureren en hanteerbaar maken, bieden modellen richting bij complexe beleidsvraagstukken. Tegelijkertijd berusten zij onvermijdelijk op keuzes, aannames, onzekerheden en vereenvoudigingen. Wanneer modeluitkomsten bestuurlijk betekenis krijgen, wordt een discussie over een modellen daarmee een discussie over beleid, publieke waarden en verantwoordelijkheid.

Een uitkomst van deze discussie is de politieke vraag naar theoretische kaders en instrumenten voor validatie, verificatie, transparantie en navolgbaarheid bij de inzet van modellen. De motie van de leden Omtzigt en Six Dijkstra vraagt onder meer naar het ontwikkelen van een wetenschappelijke standaard voor modellen, hoe afdwingbaar kan worden gemaakt dat modellen en algoritmes openbaar en navolgbaar zijn, en hoe een wettenschappelijke ‘bijsluiter’ voor modellen eruit kan zien. Tegen deze achtergrond en middels subsidie van het ministerie van Binnenlandse Zaken en Koninkrijksrelaties onderzoekt wij in dit rapport hoe verantwoord modelgebruik binnen de overheid kan worden ondersteund.

De centrale conclusie van dit onderzoek is dat de zoektocht naar één uniforme wetenschappelijke standaard voor modelgebruik weinig kansrijk is. Verschillende modellen verschillen sterk in doel, technische vorm, onzekerheid, beleidscontext, maatschappelijke impact en besluitnabijheid. Bovendien ontstaat de betekenis van modeluitkomsten niet in de berekening zelf, maar in de manier waarop zij worden geïnterpreteerd, gebruikt en verbonden met bestuurlijke keuzes. De vraag of modelgebruik verantwoord is, kan daarom

niet uitsluitend worden beantwoord aan de hand van technische kwaliteit of wetenschappelijke validatie.

Dit rapport stelt daarom legitimiteit van modelgebruik centraal. Modellen zijn niet algemeen goed of fout; uitkomsten krijgen as betekenis in gebruik en context, welke vaak politieke en normatieve dimensies heeft. De opgave voor de overheid is niet om onzekerheid expliciet, bespreekbaar en bestuurlijk hanteerbaar te maken. Legitimiteit van modelgebruik rust op een toereikende wetenschappelijke en technische basis, een expliciete bestuurlijke rechtvaardiging van de rol die modeluitkomsten krijgen, en een proces waarin verantwoordelijkheid, tegenspraak, transparantie en herbeoordeling zijn georganiseerd. Een mate van zekerheid rondom of vertrouwen in een model kan daarbij de vorm aannemen van validatie, audits, kwaliteitsborging, standaarden of certificering. Maar vertrouwen kan ook rusten op de integriteit en werkwijze van een kennisinstituut, op een robuuste organisatiepraktijk of op een proces waarin belanghebbenden en diverse perspectieven tijdig worden betrokken.

Vanuit dat perspectief introduceert het rapport een gespreksinstrument voor verantwoord modelgebruik. Dat instrument helpt betrokkenen om te bespreken en vast te leggen wanneer en waarom het beleidsgebruik van een modelmatige representatie van de wereld passend is, waarop de uitkomsten berusten, welke betekenis daaraan mag worden toegekend, welke onzekerheden en grenzen gelden, hoe verantwoordelijkheid en tegenspraak zijn georganiseerd, welke informatie nodig is voor navolgbaarheid en wanneer herbeoordeling of beëindiging nodig is. Daarmee ondersteunt het instrument een gedeelde taal voor modelontwikkelaars, beleidsmakers, bestuurders, uitvoerders, toezichthouders, volksvertegenwoordigers en - waar passend - geraakte groepen. Anders dan een ‘kwaliteitsstempel’ ondersteunt dit een overheid die modelgebruik bestuurlijk bespreekbaar maakt.

1. Modelmatig besturen

Modellen spelen een steeds grotere rol in beleid, uitvoering en evaluatie. Rekenmodellen ondersteunen de overheid bij het formuleren van klimaatdoelen, het maken van economische ramingen, het verdelen van schaarse middelen en het verkennen van toekomstige ontwikkelingen. Lange tijd werden modellen vooral toegepast in het fysieke en economische domein, maar inmiddels reiken zij verder. Zo spelen algoritmische modellen een steeds centralere rol, ook in het debat. Uitvoeringsorganisaties, toezichthouders en gemeenten gebruiken algoritmen (in verschillende varianten) en risicomodellen om aandacht, capaciteit of ondersteuning te richten.

Met die groeiende betekenis worden modellen ook vaker onderwerp van publiek en politiek debat. Een bekend voorbeeld is AERIUS in het stikstofdossier. AERIUS is een software-instrument dat (op basis van onderliggende rekenmodellen) stikstofneerslag berekent alsmede waar die neerkomt. De uitkomsten hebben grote bestuurlijke en juridische gevolgen: vergunningen kunnen worden geweigerd, projecten kunnen stilvallen en economische activiteiten kunnen worden beperkt. Juist vanwege die doorwerking keren in rechtszaken, politieke debatten en maatschappelijke discussies steeds dezelfde vragen terug: hoe werkt het onderliggende model, hoe betrouwbaar zijn de uitkomsten, welke onzekerheden kleven eraan en welke rol mogen zij spelen in besluitvorming? Van AERIUS is dan ook uitgebreide documentatie en wetenschappelijke onderbouwing beschikbaar gesteld.[1] Een ander model in het oog van het publieke debat is het strategische luchtvaartmodel AEOLUS, dat zich richt op prognoses van onder andere de uitstoot van de luchtvaart. Omdat dit over toekomstige ontwikkelingen gaat, worden daar keuzes en aannames gebaseerd op niet alleen al vastgesteld en voorgenomen beleid, maar ook op niet vast te stellen aannames zoals toekomstige uitstootreductie gebaseerd op de verwachte toekomstige efficiëntiewinsten en vlootvernieuwing.[2] Dit zijn verwachtingen en aannames over de verwachte doorontwikkeling van vliegtuigen, niet de werkelijkheid. Deze aannames, beïnvloeden via het model wel de keuzes die nu worden gemaakt. Ook bij de berekening van geluid en geluidsoverlast rond luchthavens spelen modellen een belangrijke rol. Daarbij worden gemeten geluidsgegevens van verschillende vliegtuigtypen met behulp van rekenmodellen vertaald naar de verwachte geluidsbelasting rondom een Schiphol. Hierin zitten ook aannames, zoals met betrekking tot de gemiddelde weersomstandigheden. Daarin kunnen verschillen tussen gemeten, berekende en ervaren geluid ontstaan[3]. Tot slot een voorbeeld uit een ander domein. Het CPB gebruikt een model om de

[1] https://www.aeriusproducten.nl/wetenschappelijke-onderbouwing
[2] Memo AEOLUS doorrekeningen KEV2024
[3] Tweede Kamer (2019) 31 936 (Luchtvaartbeleid) nr. 631 . Er zijn ook diverse toegankelijke uitleggen van hoe geluidsbelasting wordt bepaald beschikbaar, bijv. bij NRC.

productiviteitsontwikkeling in Nederland te meten. De publieke sector vormt een aanzienlijk onderdeel van de economie, maar vanwege de aard van die sector, wordt productiviteitsgroei moeilijk te meten en vaak gebaseerd op schattingen.[4] . Ook hier geldt dat de interpretatie van de uitkomsten niet los kan worden gezien van de gemaakte aannames en de daarmee samenhangende onzekerheden die ten grondslag liggen aan de berekeningen van het model.

Opvallend is dat niet alleen het *gebruik* van modellen kritiek oproept. Ook het ontbreken van modelmatige onderbouwing kan tot discussie leiden. In beleidsdomeinen waarin risico's, effecten of toekomstige ontwikkelingen niet direct observeerbaar zijn, rijst zonder model de vraag waarop beleidskeuzes dan berusten. Gebruikt de overheid wel een model, dan verschuift de aandacht naar de aannames, keuzes en onzekerheden die aan het model ten grondslag liggen. Modellen ontsnappen zo niet aan het politieke debat, maar worden er zelf inzet van.

Daarmee tekent zich een spanning af die kenmerkend is voor hedendaags bestuur. In domeinen als klimaat, stikstof, epidemiologie, economie, infrastructuur, sociale zekerheid en toezicht is beleid vaak afhankelijk van modelmatige representaties. Modellen spelen dan een centrale rol in de manier waarop overheden de werkelijkheid begrijpen en beleid vormgeven. Ze verschaffen een kennisbron voor beleid, en vormen soms de belangrijkste kennisbasis. De beperkingen, onzekerheden, aannames, gebruikte data, en rekenwijzen, hebben dan een groot effect op beleid. Dat kan gaan om de kennis die gebruikt wordt in beleidsvorming en de kennis en informatie die in de uitvoering van beleid wordt gebruikt. Soms gaat het nog verder en is beleid op een bepaald domein niet of nauwelijks voorstelbaar, denkbaar of uitvoerbaar, zonder de rekenmodellen, omdat de relevante werkelijkheid niet direct zichtbaar, meetbaar of bestuurlijk hanteerbaar is. In dit rapport spreken we dan over modellen die *constituerend* zijn voor beleid. Die afhankelijkheid brengt het risico met zich mee dat modeluitkomsten een status krijgen die de onderliggende aannames, onzekerheden en normatieve keuzes niet kunnen dragen.

Dat risico wordt groter wanneer modeluitkomsten in de beleidsketen verschuiven. Een scenario kan worden behandeld als voorspelling, een indicatie als bewijs, een risicosignaal als prioriteit, een kaartbeeld als bestuurlijke werkelijkheid of een wetenschappelijke uitkomst als juridische besluitgrond.[5] In die verschuiving ligt veel van de bestuurlijke gevoeligheid van modelgebruik. Niet alleen de vraag of een model de bedoelde kennis levert

[4] CPB (2017) Productiviteitsontwikkeling van de Nederlandse overheid
[5] Zie voor een analyse van de Toeslagen/CAF casus vanuit dit perspectief Giest & Klievink (2024)

is dan van belang, maar ook de vraag welke betekenis de overheid aan de uitkomst geeft en welke gevolgen daaraan worden verbonden.

Discussies over modellen zijn daarom altijd ook discussies over beleid, publieke waarden en bestuurlijke keuzes. Dat valt niet aan (enkel) de model-kant te ondervangen want dan zouden politieke en normatieve keuzes worden verscholen achter technisch ogende uitkomsten. Modellen kunnen de werkelijkheid niet volledig vervangen. Zij maken bepaalde aspecten zichtbaar, kwantificeerbaar en vergelijkbaar, maar laten andere aspecten buiten beeld.

Tegen deze achtergrond zien we de politieke vraag, meest concreet in de motie-Omtzigt/Six Dijkstra[6], naar een wetenschappelijke standaard, openbaarmaking, navolgbaarheid en een bijsluiter voor modellen en algoritmes. Die vraag is begrijpelijk. Zij vertrekt vanuit de behoefte aan controle op abstracte, soms moeilijk te doorgronden of onvoldoende openbare modellen. In dit rapport zien we de kernopgave echter breder in het organiseren van bestuurlijke navolgbaarheid: zichtbaar maken welk modelgebruik welke rol speelt, waarop vertrouwen in dat gebruik rust, waar de grenzen liggen en wanneer het gebruik van een model opnieuw ter discussie moet worden gesteld.

Dit rapport werkt die opgave langs twee lijnen uit. Eerst geven wij een bestuurskundig perspectief op modelgebruik: welke rollen vervullen modellen, hoe krijgen uitkomsten bestuurlijke betekenis en waarom is modelgebruik politiek en normatief geladen? Vervolgens introduceren wij een gespreksinstrument dat betrokkenen helpt de inzet, betekenis en begrenzing van modellen expliciet te maken, juist al in de beleidscyclus, zodat het een gesprek over beleid in plaats van louter over modellen wordt. Het instrument dient als redenerings- en afwegingsstructuur voor legitiem modelgebruik in beleid en uitvoering.

[6] Tweede Kamer (2025) 35 867 (Parlementaire enquête fraudebeleid en dienstverlening) nr. 26

2. Een bestuurskundig perspectief op modelgebruik

2.1 Verschillende rollen van modellen in beleid

Een model is, in algemene zin, een vereenvoudigde representatie van een deel van de werkelijkheid. Die representatie kan verschillende vormen aannemen: een rekenregel, statistisch model, simulatiemodel, voorspellend algoritme, machine-learningmodel of samenstelling van modellen, databronnen en interpretatiestappen. Voor het openbaar bestuur is de variatie in technieken, methoden, en complexiteit belangrijk, maar niet beslissend. Bestuurlijk relevant is vooral wat een model doet: welke kennis het produceert, welke beleidsopties het zichtbaar maakt, welke besluiten het ondersteunt en welke verantwoordelijkheid het verschuift, of impliciet of juist expliciet maakt.

Het onderscheid tussen modellen en algoritmen is daarom nuttig, maar moet niet te zwaar worden aangezet. Klassieke rekenkundige beleidsmodellen zijn vaak gebouwd rond begripsvorming, causale aannames en representaties van een beleidswerkelijkheid. Algoritmische systemen en machine-learningmodellen leggen vaker patronen, classificaties of correlaties bloot in data. Bij klassieke modellen zitten keuzes nadrukkelijk in begrippen, variabelen, relaties en scenario's; bij machine learning kunnen aannames sterker besloten liggen in trainingsdata, labels en optimalisatiedoelen. In beide gevallen is de bestuurlijke vraag vergelijkbaar: welke rol krijgt de uitkomst, welke aannames en beperkingen reizen mee en wie kan de toepassing begrijpen en beoordelen en bevragen?

Om de bestuurlijke rol zichtbaar te maken, onderscheiden wij modellen langs twee assen. De eerste as betreft de tijdsoriëntatie: beschrijft het model een bestaande of bekende situatie, of probeert het toekomstige ontwikkelingen te voorspellen of te verkennen? De tweede as betreft de bestuurlijke inzet: wordt het model vooral gebruikt om kennis te ontwikkelen, of stuurt het directer op besluiten, verdelingen of interventies?

Tijdsoriëntatie / rol	Kennen	Besluiten
Beschrijven	**Begrijpen**: luchtkwaliteitsmonitoring, SCP-inkomensanalyse, epidemiologische duiding.	**Verdelen**: verdeelmodellen, budgetverdeling, objectieve verdeelsleutels.

Voorspellen / verkennen	**Anticiperen**: CPB-ramingen, PBL-klimaatverkenningen, scenarioanalyse.	**Sturen**: AERIUS, risicoselectie, voorspellende handhaving en toezicht.

Tabel 1: typologie

Deze typologie is geen harde classificatie. Hetzelfde model kan meerdere rollen tegelijk vervullen. Juist dat maakt modelgebruik bestuurlijk gevoelig. AERIUS (en onderliggende modellen) kan helpen om stikstofdepositie te begrijpen, maar wordt ook gebruikt in vergunningverlening. Een economisch model kan beleidsscenario's verkennen, maar ook budgettaire ruimte legitimeren. Een risicomodel kan intern signaleren, maar in de praktijk de toegang tot controle, ondersteuning of sancties mede bepalen. Wanneer een model meerdere rollen krijgt, wordt minder duidelijk waar de berekening ophoudt en het bestuurlijke oordeel begint, en moet expliciet worden gemaakt of de onderbouwing voor de ene rol ook voldoende is voor de andere. Die andere rollen kunnen op verschillende plekken in een beleidsketen zitten, maar het komt ook voor dat modellen in een andere rol, in ander beleid worden ingezet.

Daarmee komt de bestuurlijke kernvraag in beeld: niet of een model in abstracte zin goed is, maar of het adequaat is voor het doel waarvoor het in deze context wordt gebruikt. Een model kan wetenschappelijk zorgvuldig ontwikkeld zijn en toch problematisch worden gebruikt wanneer de uitkomst een zwaardere bestuurlijke, juridische of politieke betekenis krijgt dan de onderliggende aannames en onzekerheden toelaten. Omgekeerd kan een model met bekende beperkingen verantwoord worden gebruikt wanneer de rol bescheiden is, de grenzen expliciet zijn en aanvullende kennisbronnen worden meegewogen.

2.2 Hoe modeluitkomsten bestuurlijke betekenis krijgen

Een modeluitkomst spreekt niet voor zichzelf. Betekenis ontstaat in de manier waarop een model is ontwikkeld en hoe modeluitkomsten worden geïnterpreteerd en gebruikt. Een getal, score, kaartbeeld of scenario wordt geïnterpreteerd door modelleurs, vertaald door beleidsmakers, aangehaald in Kamerbrieven, gebruikt in bestuurlijke afwegingen, ingebracht in rechtszaken of toegepast in uitvoeringsbesluiten. In dat proces kan de status van de uitkomst verschuiven. Wat begon als scenario kan verwachting worden; wat bedoeld was als signaal kan prioritering worden; wat een indicatie was, kan bewijs worden.

Modelgebruik moet daarom worden begrepen als beleidspraktijk. Modellen brengen niet alleen informatie in beleid, maar structureren ook welke informatie als relevant, urgent of vergelijkbaar wordt gezien. Dat hangt samen met de logica van modelleren. Modellen werken met gegevens die beschikbaar en kwantificeerbaar zijn – met gegevens, aannames, en observaties die 'kwantificeerbaar zijn'. Kwantificering is hiermee noodzakelijkerwijs een reductie van de werkelijkheid. Wie kwantificeert, maakt vergelijkingen mogelijk en maakt sommige waarden zichtbaar. Tegelijkertijd kunnen waarden die moeilijker meetbaar zijn - zoals vertrouwen, rechtvaardigheid, ervaren ongelijkheid of bestuurlijke uitvoerbaarheid - naar de achtergrond verdwijnen. Zo beïnvloeden modellen welke aspecten van een beleidsvraagstuk zichtbaar worden, waar bestuurlijke aandacht naartoe gaat en wat uiteindelijk als beleidsmatig relevant wordt beschouwd (Stone, 2012; Vydra & Klievink, 2019).

Verantwoord modelgebruik vraagt daarom niet alleen de vraag of het model valide is, maar ook welke waarden en effecten wel en niet in het model zijn opgenomen. Welke aannames over gedrag, ruimte, economie, risico of naleving zijn gemaakt? Welke informatie ontbreekt omdat zij niet goed kwantificeerbaar is? Welke distributieve gevolgen kunnen ontstaan wanneer modeluitkomsten leidend worden? En welke politieke keuzes dreigen te worden gedelegeerd aan experts of aan een modelmatige representatie?

Bij het gebruik van modellen in beleid, uitvoering en toezicht, kan de grens tussen wat bekend is en wat onzeker of speculatief blijft vervagen. In veel domeinen van overheidsbeleid, inclusief de uitvoering ervan en toezicht erop, zijn er puzzelstukjes waarover wetenschappelijk veel zekerheid bestaat en het model zorgvuldig is ontwikkeld, terwijl de beleidsvraag betrekking heeft op de hele puzzel. Het model kan dan de indruk wekken dat ook die hele puzzel in dezelfde mate kenbaar is. Een model kan echter haast per definitie niet de werkelijkheid volledig of definitief weergeven. Betekenis en bruikbaarheid van modeluitkomsten zijn altijd afhankelijk van hun doel en gebruikscontext (Oreskes et al., 1994; Parker, 2018; Edmonds et al., 2019). Het draait daarom om de vraag welke uitspraken het model verantwoord kan dragen binnen een specifieke context. Juist daar is bestuurlijke navolgbaarheid nodig: onzekerheid en gevoeligheid zijn kenmerken van modellen, daarom is het van belang zichtbaar te maken waar kennis sterk is, welke aannames (on)problematisch zijn en waar politieke oordeelsvorming begint.

Daarbij komt dat niet modelleren óók een keuze is en dat niet modelleren niet altijd mogelijk is.[7] In domeinen als stikstof, klimaat, epidemiologie of economie is de

[7] "*Idealiter zou je een soort klinische studie doen naar een bevolking met een avondklok en een zonder, maar dat kon niet*", zegt Jacco Wallinga (RIVM) bij verhoor parlementaire enquêtecommissie Corona. (NOS, 15-6-2026)

beleidswerkelijkheid vaak niet direct observeerbaar. Zonder modellen is er dan geen stabiele kennisbasis om beleid op te baseren. In zulke gevallen zijn modellen niet alleen ondersteunend, maar constituerend of epistemisch: zij vormen de kennisinfrastructuur die mede zichtbaar maakt wat het beleidsprobleem is, welke variabelen ertoe doen en welke handelingsruimte denkbaar wordt. Soms is 'de werkelijkheid' nauwelijks anders meetbaar, of althans niet binnen redelijke besteding van publieke middelen (ook een publieke waarde). Dat maakt modelgebruik onvermijdelijk, maar ook politiek-bestuurlijk zwaarwegend.

2.3 Modelgebruik is politiek

Modellen zien we daarom als socio-technische kennisinstrumenten. Modellen combineren wetenschappelijke inzichten, technische keuzes, databronnen, organisatorische routines en bestuurlijke doelen. Ook wanneer een model methodologisch valide is, bevat het keuzes over wat meetelt, wat buiten beeld blijft, welke onzekerheden aanvaardbaar worden geacht en welke rol de uitkomst krijgt. Die keuzes worden soms technisch gemaakt, maar behoren evenzeer tot het sociale. Zij raken aan publieke waarden, rechtvaardigheid, rechtsbescherming, uitvoerbaarheid en democratische controle.

Modelgebruik heeft onvermijdelijk een bestuurlijke en politieke dimensie. Modeluitkomsten krijgen pas betekenis binnen een beleidscontext waarin keuzes moeten worden gemaakt over doelen, belangen en waarden. Een model levert informatie, maar bepaalt niet hoe die informatie moet worden gewogen, welke uitkomsten doorslaggevend zijn of welke gevolgen maatschappelijk en bestuurlijk aanvaardbaar zijn. Die afwegingen zijn onvermijdelijk politiek en behoren tot de verantwoordelijkheid van bestuurders en beleidsmakers.

Het gespreksinstrument dat wij voorstellen richt zich overigens niet primair op de politieke besluitvorming zelf, maar op het beleidsmatige werk waarin politieke wensen worden vertaald naar uitvoerbaar en verantwoord beleid. Juist in die fase is een risico op een problematische prikkel: beleid wordt vormgegeven alsof de benodigde kennisbasis al beschikbaar is, waarna van modellen en modeleigenaren wordt gevraagd om te gaan onderbouwen dat zij die kennis kunnen leveren. Wanneer dat niet goed lukt, wordt de oplossing al snel gezocht in verdere modelontwikkeling.[8] Daarmee verschuift de vraag naar de modelkant, terwijl de kernvraag breder is: welke kennis over de relevante werkelijkheid is nodig om dit beleid verantwoord te kunnen ontwerpen, uitvoeren en verantwoorden, en kunnen bestaande modellen die werkelijkheid voldoende representeren? Beleidsmakers moeten daarom toetsen of de daarvoor noodzakelijke kennisbasis bestaat, welke rol

[8] Enkele respondenten merkten op dat budgetten voor modelontwikkeling beperkt zijn en vroegen zich af of de verhouding tussen de investering in modellen en de miljardenbeslissingen die erop leunen, wel in verhouding zijn.

modellen daarin spelen en waar de grenzen van die modellen liggen. De uitkomst daarvan zou kunnen zijn dat een beleidsvoornemen niet zonder meer mogelijk is, of alleen onder duidelijke voorwaarden, onzekerheden en beperkingen. In dat geval moet de spanning tussen beleidsdoel, benodigde kennis en beschikbare modelrepresentatie niet uitsluitend worden opgelost binnen modelontwikkeling of uitvoering, maar onderdeel worden van de beleids- en politiek-bestuurlijke afweging: accepteren we de beperkingen, passen we het beleid aan, investeren we in kennis- of modelontwikkeling, of achten we het beleid in deze vorm niet uitvoerbaar? De leidraad beoogt beleidsmakers te helpen om die afweging eerder, explicieter en beter verantwoord te maken en daarmee scherper te krijgen of de vervolgvraag inderdaad bij de modellen kan worden gelegd, of terug zou moeten worden gelegd bij politieke besluitvorming.

Deze politieke dimensie betekent niet dat de ontwikkeling van modellen partijdig of onwetenschappelijk is. Integendeel, wetenschappelijke kwaliteit is een noodzakelijke voorwaarde voor verantwoord modelgebruik. Die kwaliteit bepaalt echter niet welke politieke betekenis aan een modeluitkomst mag worden toegekend. Een model kan geschikt zijn om een landelijke trend te analyseren, maar niet om individuele besluiten te rechtvaardigen. Het kan scenario's inzichtelijk maken, maar niet bepalen welke verdeling van middelen of risico's rechtvaardig is. En een algoritme kan risicopatronen signaleren, maar niet zelfstandig beslissen wie welke behandeling, voorziening of maatregel behoort te krijgen.

Het publieke debat over modellen kan daarom worden begrepen als een noodzakelijke vorm van herpolitisering. Het brengt de fundamentele keuzes rond modelontwikkeling en modelgebruik – keuzes die lange tijd vooral als technische, wetenschappelijke of uitvoeringsvraagstukken zijn behandeld – terug naar de plek waar zij thuishoren: de beleids- en politiek-bestuurlijke arena. Dat is van belang omdat beleid en politiek de neiging kan hebben normatieve afwegingen te laten opgaan in het werk van modelontwikkelaars, kennisinstituten of uitvoeringssystemen. Dat kan bestuurlijk aantrekkelijk zijn: een model of modeluitkomst wekt de indruk van objectiviteit en biedt ogenschijnlijk houvast. Maar modeluitkomsten zijn niet objectief, er liggen keuzes aan ten grondslag die niet mogen verdwijnen in een technocratisch debat over aannames, parameterkeuzes, dataselecties of beslisregels. Zij moeten expliciet worden gemaakt en democratisch kunnen worden besproken en verantwoord aangezien de legitimiteit van een model niet uitsluitend een technische eigenschap is, maar hangt deze ook af van de

kwaliteit van de kennisbasis, de explicitering van onzekerheden en aannames, en de wijze waarop relevante perspectieven in de besluitvorming worden betrokken en afgewogen.

2.4 Van standaard naar legitimiteit, assurance en vertrouwen

De politieke vraag naar een wetenschappelijke standaard is begrijpelijk, maar te smal. Zij suggereert dat kan worden vastgesteld of een model goed genoeg is door het model te toetsen aan algemene wetenschappelijke criteria. Voor onderdelen van modelontwikkeling kan dat zinvol zijn. Er bestaan good modelling practices, onzekerheidsstandaarden, validatiepraktijken, modeldocumentatie, algoritmische transparantiestandaarden, impact assessments en – met name voor AI - toenemende wettelijke en normatieve kaders. Die instrumenten blijven belangrijk. Zij geven vormen van *assurance*: redenen om te vertrouwen of te kunnen controleren dat een model, systeem of organisatie aan bepaalde eisen voldoet.

Maar assurance is niet hetzelfde als legitimiteit. Legitiem modelgebruik vraagt dat duidelijk is waarom dit model of deze modelconfiguratie in deze beleidscontext passend is, welke rol de uitkomst mag krijgen, welke belangen worden geraakt, welke onzekerheden blijven bestaan en hoe tegenspraak en herbeoordeling zijn georganiseerd. Soms rust legitimiteit vooral op wetenschappelijke validatie; soms op vertrouwen in de expertise van een kennisinstituut; soms op externe audit of toezicht; soms op participatie en publieke betwistbaarheid. Vaak is een combinatie nodig.

Het begrip vertrouwensgrond helpt om dat zichtbaar te maken. Vertrouwen heeft daarbij niet uitsluitend betrekking op het model zelf, maar op de mensen, organisaties en processen die het model ontwikkelen, toepassen, interpreteren en gebruiken in de besluitvorming. Waarop rust het vertrouwen dat modelgebruik verantwoord is? Op de integriteit van modelleurs? Op een gedocumenteerde werkwijze? Op validatie en onzekerheidsanalyse? Op een kwaliteitsborgingssysteem? Op wettelijke eisen of standaarden? Op institutioneel toezicht? Op betrokkenheid van belanghebbenden? Of op de mogelijkheid om uitkomsten te betwisten en gebruik te stoppen? Door die vertrouwensgrond expliciet te maken, wordt zichtbaar wanneer aanvullende waarborgen nodig zijn.

Die afweging is niet te vangen in een standaard of checklist. Dat kan zelfs averechts werken wanneer het de indruk wekt dat naleving van een format of bijsluiter voldoende is. Dan

ontstaat compliance theater: de papieren werkelijkheid voldoet, terwijl de betrokken actoren het gesprek over passend gebruik, waarden, gevolgen en verantwoordelijkheid niet werkelijk is gevoerd. Een zorgvuldig gesprek is de kern maar ook vluchtig. Wanneer de gemaakte afwegingen niet worden vastgelegd, zijn zij later moeilijk navolgbaar, toetsbaar of opnieuw te beoordelen. Verantwoord modelgebruik vraagt daarom ook om een vorm van institutioneel geheugen waarin de gemaakte keuzes, aannames, verantwoordelijkheden en waarborgen worden vastgelegd. Een optie hiervoor is de vastlegging van de gespreksuitkomsten in een bijlage bij bijvoorbeeld het Beleidskompas, of in een ander dossier. In dit rapport stellen we het gesprek zelf centraal en schrijven geen specifieke vorm van vastlegging voor. Vastlegging moet hoe dan ook niet het einde van het gesprek zijn, maar het zichtbare spoor van een gesprek dat daadwerkelijk heeft plaatsgevonden.

3. Zes bestuurlijke afwegingen bij modelgebruik

Het gespreksinstrument is ontwikkeld op basis van literatuuronderzoek, analyse van internationale standaarden en instrumenten, gesprekken met experts en bijeenkomsten met een adviesgroep. Daaruit komt een dubbel beeld naar voren. Enerzijds verschillen modellen sterk in doel, opzet, domein en toepassing. Anderzijds keren in vrijwel alle modelpraktijken vergelijkbare bestuurlijke vragen terug: waarom wordt modellering ingezet, waarop berusten uitkomsten, welke betekenis mogen zij krijgen, wie draagt verantwoordelijkheid, welke informatie is nodig en wanneer moet opnieuw worden beoordeeld?

Anders dan een technische standaard, keurmerk of afvinklijst – hoe comfortabel een dergelijk instrument ook zou zijn – is het gespreksinstrument een gestructureerde redeneringsroute die de 'assurance' vraag terug bij het beleid en indien nodig bij de politiek legt. De zes stappen in het instrument brengen steeds twee werelden bij elkaar. Aan de ene kant staat wat modelleurs, data-experts, kennisinstituten of leveranciers kunnen bieden: representaties, data, aannames, validatie, onzekerheidsanalyse, documentatie en kwaliteitsborging. Aan de andere kant staat wat beleid vraagt: een antwoord op een publieke vraag, een keuze tussen opties, een verdeling van middelen, een uitvoeringshandeling, een juridische onderbouwing of een publieke rechtvaardiging. Verantwoord modelgebruik ontstaat wanneer expliciet wordt gemaakt of wat de modelwereld kan bieden voldoende is voor wat de beleidswereld ervan vraagt.

Daarmee fungeert het instrument als boundary object.[9] Het ondersteunt verschillende betrokkenen om over hetzelfde modelgebruik te spreken zonder te doen alsof zij dezelfde expertise, verantwoordelijkheid of informatiebehoefte hebben. Het gesprek is niet vrijblijvend: het moet leiden tot vastlegging van rolclaim, kennisbasis, vertrouwensgrond, waarborgen, informatiebehoeften en herbeoordelingscriteria. Tegelijkertijd mag het niet verworden tot een invuloefening. Een vraag die niet goed kan worden beantwoord, is juist een belangrijk signaal dat modelgebruik nog niet bestuurlijk navolgbaar is.

Stap	Kernvraag	Modelwereld	Beleidswereld
1	Is modellering passend?	Wat kan modellering representeren en waar liggen de grenzen?	Waarom is modellering nodig voor deze beleidsopgave en wie moet aan tafel?

[9] Hier gezien als hulpmiddel dat verschillende betrokkenen ondersteunt om over complexe vraagstukken met elkaar in gesprek te gaan en tot een gedeeld begrip te komen (Bell et al., 2026; Star, 2010).

2	Waarop rust de uitkomst?	Welke data, aannames, methoden, validatie en kwaliteitsborging liggen eraan ten grondslag?	Waarop rust vertrouwen in deze kennisbasis en welke afhankelijkheden bestaan?
3	Welke betekenis mag de uitkomst krijgen?	Welke claim kan de uitkomst dragen en welke onzekerheden blijven bestaan?	Mag de uitkomst dienen als signaal, scenario, onderbouwing, prioritering of besluitgrond?
4	Hoe is bestuurbaarheid georganiseerd?	Welke fouten, drift, beperkingen of technische afhankelijkheden kunnen optreden?	Wie is verantwoordelijk, wie kan tegenspreken, corrigeren, escaleren of stoppen?
5	Wat moet navolgbaar zijn?	Welke documentatie en uitleg zijn beschikbaar?	Welke informatie hebben bestuur, Kamer, toezicht, uitvoering en geraakte groepen nodig?
6	Wanneer herbeoordelen of stoppen?	Welke veranderingen in data, model, kennis of techniek zijn relevant?	Wanneer wordt gebruik aangepast, beperkt, heropend of beëindigd?

Tabel 2: gespreksstappen

3.1 Stap 1: Is modellering een passende manier om het beleidsvraagstuk te ondersteunen?

De eerste stap begint niet bij het model, maar bij de beleidsopgave. Welk probleem moet worden begrepen, welke beslissing moet worden voorbereid of welke uitvoeringshandeling moet worden ondersteund (Hall et al., 2014)? Pas daarna kan worden bepaald of modellering daarvoor passend is: dat is ook een afweging (Edmonds, 2017). Dat is van belang omdat modellen niet alleen informatie leveren, maar ook bepalen welke aspecten van de werkelijkheid zichtbaar en vergelijkbaar worden.

Een verantwoord gebruik van modellen begint bij een heldere beleidsvraag. Modelleurs kunnen aangeven wat een model kan representeren, welke relaties het plausibel kan beschrijven en waar de grenzen liggen. Beleidsmakers moeten expliciet maken wat zij nodig hebben: een diagnose, scenario, verdeling, prioritering, juridische onderbouwing of beleidskeuze. Wanneer die beleidsvraag niet duidelijk is, kan ook niet worden vastgesteld of modellering passend is.

Deze stap vraagt ook wie aan tafel moet zitten. Modelontwikkelaars zijn vaak niet dezelfde personen als modelbeheerders, softwareontwikkelaars, data-eigenaren, beleidsgebruikers of uitvoerders. Bij modelketens kunnen meerdere organisaties betrokken zijn. Waar modeluitkomsten burgers, bedrijven of maatschappelijke groepen raken, kan ook betrokkenheid van geraakte perspectieven nodig zijn. Vraag 0 van het instrument luidt daarom: wie moeten deelnemen aan dit gesprek om te voorkomen dat cruciale aannames onbesproken blijven?

Een belangrijke uitkomst van deze stap kan ook zijn dat modellering niet passend is, of alleen als verkenning. Soms zijn praktijkkennis, participatie, experimenten, juridische analyse of expliciete politieke keuzevorming belangrijker dan verdere kwantificering. Of zijn de beperkingen van modelinzet te groot om dat verantwoord te kunnen doen. Verantwoord modelgebruik omvat dus ook het recht om nee te zeggen tegen modelgebruik of om het gebruik te beperken tot scenario's.

Die keuze kan ook verder strekken: zeker wanneer modellen en algoritmen een constitutieve rol spelen in beleid. In die rol gebruikt beleid de modellen niet alleen ter ondersteuning, maar is zonder deze modellen moeilijk denkbaar of uitvoerbaar. Zij vormen een onmisbaar onderdeel van de kennisbasis waarop beleid rust – en dus moeten de consequenties daarvan al vroeg in de beleidsketen worden bekeken. Een keuze om niet op een model te bouwen, kan hier betekenen dat het beleid niet realistisch uitvoerbaar is. Tevens moet hier worden onderkend of in de voorziene beleidsopties het modelgebruik niet al besloten ligt.

3.2 Stap 2: Op welke kennisbasis zijn de uitkomsten gebaseerd?

De tweede stap richt zich op de kennisbasis van modeluitkomsten: data, aannames, rekenregels, deskundigenoordelen, modelkoppelingen, validatie, onzekerheidsanalyse en organisatorische kwaliteitsborging. Deze stap maakt zichtbaar waarop de uitkomst rust en welke keuzes daarin besloten liggen.

Modellen zijn selectieve representaties van de werkelijkheid. Zij maken sommige factoren zichtbaar en laten andere buiten beeld (Oreskes et al., 1994; Parker, 2018). Gegevens worden verzameld, geselecteerd en gecombineerd; variabelen worden gekozen; aannames over gedrag, causaliteit, schaal en tijd worden gemaakt. Bij algoritmische systemen kunnen

aannames verscholen liggen in trainingsdata, labels, optimalisatiefuncties of historische patronen. Achter gegevens gaan keuzes schuil die worden beïnvloed door wat meetbaar is en door wat wordt gemeten, vastgelegd en gebruikt. Deze keuzes maken onderdeel uit van de kennisbasis waarop modeluitkomsten rusten (Grimm et al., 2014).

Aannames zijn ook een belangrijk onderdeel van de kennisbasis van een model. Modellen bevatten aannames over bijvoorbeeld menselijk gedrag, economische ontwikkelingen, maatschappelijke trends, beleidsreacties of de werking van uitvoeringsprocessen. Dergelijke aannames maken het mogelijk om complexe vraagstukken hanteerbaar te maken. Tegelijkertijd beïnvloeden zij de uitkomsten die een model produceert (Kloprogge et al., 2011). Verantwoord modelgebruik vraagt daarom om inzicht in de aannames die worden gemaakt, de onderbouwing daarvan en de mate waarin zij aansluiten bij het vraagstuk waarvoor het model wordt ingezet (Funtowicz & Ravetz, 1990; van der Sluijs et al., 2005).

Modeluitkomsten komen zelden voort uit één geïsoleerd rekenmodel. In de praktijk wordt vaak gebruikgemaakt van combinaties van modellen, databronnen en menselijke beoordelingen. Modellen worden gekoppeld aan andere modellen, gevoed met gegevens uit verschillende systemen of aangevuld met deskundigenoordelen. In de literatuur wordt daarom steeds vaker gesproken over modelketens. In dergelijke modelketens kunnen aannames, onzekerheden en beperkingen zich door meerdere modellen en databronnen heen verplaatsen (Filatova et al., 2025; Schröter et al., 2025). De kennisbasis van modeluitkomsten strekt zich in dergelijke situaties uit over de gehele modelketen. Het gesprek over de onderbouwing van modeluitkomsten vraagt daarom om aandacht voor de gegevens, aannames en beoordelingen die binnen die keten een rol spelen.

Voor modelleurs gaat deze stap over het leveren van een zo helder mogelijke onderbouwing: welke bronnen zijn gebruikt, welke kwaliteitscontroles zijn uitgevoerd, welke onzekerheden zijn onderzocht, welke standaarden of good modelling practices zijn gevolgd en welke beperkingen blijven bestaan. Voor beleid gaat deze stap over de vertrouwensgrond: waarom mogen wij op deze kennisbasis vertrouwen voor deze beleidsopgave?

Die vertrouwensgrond kan technisch, organisatorisch, institutioneel of procedureel zijn. Validatie en verificatie zijn belangrijk, maar niet altijd voldoende. Soms is vertrouwen mede gebaseerd op de integriteit en onafhankelijkheid van een kennisinstituut, op peer review, op een gedragscode wetenschappelijke integriteit, op externe audit, op domeinstandaarden, op NEN- of ISO-processen, op wettelijke eisen of op publiek toegankelijke documentatie.

De focus van het gespreksinstrument ligt op het model en diens aard in relatie tot de vormgeving en uitvoerbaarheid van het beleid waarvoor het model nodig wordt geacht. Er zijn voor verschillende type modellen of aspecten van inzet, al veel wettelijke en normatieve kaders, en in sommige gevallen ook standaarden. Wanneer het gaat om AI-modellen, is er Europese wetgeving en wordt er gewerkt aan standaarden. Wanneer het gaat om algoritmes met impact op mensenrechten, is er in Nederland de IAMA. Wanneer er privacyaspecten spelen, is een DPIA nodig. Maar niet altijd valt modelgebruik daar onderen niet altijd levert het doorlopen van dergelijke instrumenten of voldoen aan normen of eisen, een integrale afweging op. In het gespreksinstrument komt daarom de vraag naar voren welke van deze instrumenten van toepassing zijn en helpt om te bepalen wanneer zij relevant zijn en welke vragen zij niet afdekken.

Deze stap maakt daarmee expliciet welke vorm van assurance in een specifiek model-gebruik-beleid scenario relevant is.

3.3 Stap 3: Welke betekenis kan aan de uitkomsten worden toegekend?

De derde stap is het analytische hart van het instrument. Zij vraagt welke rolclaim aan de modeluitkomst mag worden verbonden. Een modeluitkomst kan een signaal zijn, een indicatie, een scenario, een risico-inschatting, een prioritering, een onderbouwing of een besluitgrond. Die rollen verschillen in bestuurlijk gewicht en vragen om verschillende vormen van bewijs, uitleg en waarborg.

Deze stap voorkomt dat technische kwaliteit wordt verward met bestuurlijke draagkracht. Een model kan robuust zijn voor een bepaald doel, maar ongeschikt voor een ander doel. Het AERIUS-voorbeeld laat dit scherp zien: een rekenmodel kan waardevol zijn voor gebiedsgerichte of landelijke analyse, terwijl gebruik voor individuele vergunningverlening andere eisen stelt aan schaalniveau, onzekerheid, bewijswaarde en juridische doorwerking. Het probleem ontstaat niet alleen in het model, maar in de verschuiving van de rol die de uitkomst krijgt.

Daarom moet expliciet worden vastgesteld welke conclusies op basis van het model wel en niet verantwoord zijn. Welke onzekerheden zijn gekwantificeerd? Welke niet? Welke waarden of effecten zijn buiten beeld gebleven? Welke aanvullende kennis is nodig? Waar

moet politieke of bestuurlijke oordeelsvorming expliciet plaatsvinden? Deze stap ondersteunt ook het bepalen van een verwerpingscriterium: onder welke condities is de uitkomst niet meer bruikbaar voor de beoogde rol?

De in paragraaf 2.1 geïntroduceerde 2×2-typologie biedt hiervoor een denkkader. De positie van een model in de typologie is medebepalend voor de eisen die aan de onderbouwing, validatie en omgang met onzekerheden mogen worden gesteld. Modellen die primair worden ingezet om te sturen op besluiten, verdelingen of interventies vragen om een zwaardere onderbouwing dan modellen die vooral worden gebruikt om te begrijpen of te anticiperen. De eisen aan validatie, nauwkeurigheid en robuustheid moeten daarom aansluiten bij de bestuurlijke rol die het model vervult. Een scherp geformuleerde rolclaim helpt zowel modelleurs als beleidsmakers. Modelleurs hoeven niet meer te suggereren dat een model meer kan dan het kan dragen. Beleidsmakers krijgen duidelijker in beeld wat zij zelf moeten beoordelen, afwegen en verantwoorden. Zo wordt voorkomen dat normatieve keuzes ongemerkt worden uitbesteed aan modeltechniek.

Dit is in het bijzonder van belang wanneer modellen constituerend zijn voor beleid. In dat geval is beleid belangrijke mate model-afhankelijk: zonder het model bestaat er geen of maar beperkt direct observeerbare informatie waarop beleidskeuzes kunnen worden gebaseerd

3.4 Stap 4: Hoe zijn verantwoordelijkheid, menselijke tussenkomst en tegenspraak georganiseerd?

De vierde stap richt zich op bestuurbaarheid. Modellen dragen zelf geen verantwoordelijkheid; mensen en organisaties doen dat. Daarom moet duidelijk zijn wie opdrachtgever, eigenaar, beheerder, gebruiker, besluitvormer en toezichthouder is, en wie kan ingrijpen wanneer modelgebruik problematisch wordt.

Dit is des te belangrijker omdat modelpraktijken steeds vaker gespecialiseerd en ketengewijs zijn. De modelbouwer, softwareontwikkelaar, datahouder, beleidsafdeling, uitvoerder, leverancier en toezichthouder kunnen verschillende organisaties of teams zijn. Zonder expliciete verantwoordelijkheidsverdeling ontstaat een many-hands problem: iedereen is betrokken, maar niemand is volledig aanspreekbaar.

Menselijke tussenkomst is alleen betekenisvol wanneer mensen tijdig, geïnformeerd en bevoegd kunnen oordelen. Een *human in the loop* is onvoldoende als de menselijke rol feitelijk neerkomt op het bevestigen van modeluitkomsten zonder ruimte, kennis of mandaat om af te wijken. Deze stap vraagt daarom naar beoordelingsruimte, escalatieroutes, correctiemogelijkheden, fallbackprocedures en stopbevoegdheid.

Tegenspraak hoort bij legitimiteit. Professionele expertise, uitvoeringskennis, juridische toetsing, maatschappelijke signalen en perspectieven van geraakte groepen kunnen laten zien dat een modeluitkomst niet past bij de werkelijkheid of dat het gebruik ongewenste gevolgen heeft. Het instrument vraagt daarom niet alleen wie verantwoordelijk is, maar ook hoe georganiseerde tegenspraak mogelijk blijft.

3.5 Stap 5: Welke informatie is nodig om het gebruik te begrijpen en te beoordelen?

De vijfde stap richt zich op navolgbaarheid en transparantie. Transparantie betekent niet dat alle code, data en documentatie voor iedereen op dezelfde manier openbaar moeten zijn. Het betekent dat betrokkenen de informatie krijgen die zij nodig hebben om modelgebruik te begrijpen, beoordelen, controleren of betwisten.

Die informatiebehoefte verschilt per actor. Modelleurs hebben technische documentatie nodig. Beleidsmakers moeten weten welke rol de uitkomst mag spelen en welke onzekerheden relevant zijn. Bestuurders moeten zicht hebben op risico's, verantwoordelijkheden en politieke gevolgen. Toezichthouders moeten kunnen reconstrueren hoe het model is gebruikt. Volksvertegenwoordigers moeten democratische controle kunnen uitoefenen. Burgers of bedrijven die geraakt worden, moeten kunnen begrijpen welke rol modelgebruik speelt en hoe zij fouten of gevolgen kunnen betwisten.

Bij deze stap hoort het bepalen van welke vorm van vastlegging en communicatie, en jegens welk publiek, passend is. Goede vastlegging of communicatie maakt duidelijk waarvoor het model wel en niet bedoeld is, welke rol uitkomsten krijgen, welke onzekerheden en beperkingen gelden, wie verantwoordelijk is, welke aanvullende kennis wordt gebruikt en eventueel hoe vragen, klachten of betwisting mogelijk zijn. Op die manier kan voorkomen worden dat in toekomst een beeld of verwijt ontstaat dat modellen zijn ingezet zonder dat de juiste vragen zijn gesteld. Voor de goede orde: ook nu is er veel kennis, worden er goede

vragen gesteld en kritisch naar modelgebruik gekeken. Maar de timing en plek in het beleidsproces speelt een belangrijke rol om te voorkomen dat die vragen een politiek of bestuurlijk momentum missen, of als ‘vragen achteraf’ worden gezien.

Een belangrijke waarschuwing is dat documentatie ervan, het gesprek niet mag vervangen. Een verslag of dossier kan modelgebruik niet legitimeren; zij kan hoogstens zichtbaar maken hoe legitimiteit is onderbouwd en waar grenzen liggen.

3.6 Stap 6: Wanneer beoordelen we opnieuw of het model nog passend is?

De zesde stap maakt duidelijk dat verantwoord modelgebruik een proces is. Een model kan op enig moment passend zijn, maar later niet meer. Beleidsdoelen veranderen, wetgeving wijzigt, data verouderen, maatschappelijke omstandigheden verschuiven, nieuwe kennis ontstaat, modelversies veranderen en gebruik kan zwaarder worden dan oorspronkelijk bedoeld.

Daarom moet vooraf worden vastgelegd wanneer herbeoordeling nodig is. Aanleiding kan bijvoorbeeld zijn dat de bestuurlijke rol van het model verandert, zoals wanneer een model dat oorspronkelijk bedoeld was om ontwikkelingen te anticiperen in de praktijk steeds vaker wordt gebruikt om besluiten te sturen. Andere relevante triggers zijn toepassing in een nieuwe context, koppeling aan andere modellen, wijziging van data of definities, nieuwe wetenschappelijke inzichten, signalen uit uitvoering, maatschappelijke controverse, juridische uitspraken, onverklaarbare afwijkingen tussen model en praktijk of aanwijzingen voor ongelijke effecten tussen groepen.

Herbeoordeling moet meer zijn dan periodieke actualisatie. Soms moet een model worden aangepast, beperkt, aangevuld met andere kennis of tijdelijk stopgezet. Soms moet het gebruik fundamenteel worden herzien omdat de beleidsvraag inmiddels iets anders vraagt dan het model kan bieden. Daarmee introduceert deze stap stoplogica: criteria voor aanpassing, beperking, pauzering of beëindiging van modelgebruik.

De cyclus sluit aan bij het idee dat de zes stappen niet eenmalig zijn. Juist bij langdurig modelgebruik, modelketens en beleidsdomeinen met grote publieke gevolgen moet het

gesprek opnieuw kunnen worden geopend. Zo blijft zichtbaar of de oorspronkelijke rolclaim, kennisbasis, vertrouwensgrond en waarborgen nog passen bij het actuele gebruik.

3.7 De zes afwegingen in samenhang

De zes stappen hebben een cyclisch karakter. Zij beginnen bij de beleidsopgave, maken de modelbijdrage en rolclaim expliciet, organiseren bestuurbaarheid en navolgbaarheid, en eindigen met de vraag wanneer het gesprek opnieuw moet worden gevoerd. Daarmee ondersteunen zij wijsheid in modeltoepassing: niet meer vertrouwen vragen van een model dan het kan dragen, niet minder gebruikmaken van modelkennis dan verantwoord mogelijk is, en steeds zichtbaar houden waar politieke oordeelsvorming nodig blijft.

4. Het gespreksinstrument in de praktijk

4.1 Wie praten er mee?

Het instrument is bedoeld om werelden met elkaar in gesprek te brengen die in de praktijk vaak gescheiden zijn. Modelontwikkelaars, data-experts, softwareontwikkelaars, beleidsmakers, uitvoerders, juristen, bestuurders, toezichthouders en volksvertegenwoordigers kijken naar hetzelfde modelgebruik, maar stellen andere vragen. Juist daarom is het eerste praktische vraagstuk wie aan tafel moet zitten.

Wie nodig is, hangt af van de casus. Bij een intern verkennend model wat bedoeld is om de werkelijkheid te begrijpen kan soms een beperkt gesprek volstaan. Bij modelgebruik met juridische, distributieve of maatschappelijke gevolgen – een model dat 'stuurt' – is een breder gesprek nodig. Dan gaat het niet alleen om technische expertise, maar ook om beleidsinhoud, uitvoering, juridische toetsing, publieke waarden en, waar passend en mogelijk, perspectieven van groepen die door modeluitkomsten worden geraakt.

Het gespreksinstrument brengt perspectieven samen, maar mag verantwoordelijkheid niet laten verdampen. Ook na een goed gesprek moet duidelijk blijven wie eigenaar is van het modelgebruik, wie de rolclaim bewaakt, wie wijzigingen signaleert, wie betrokkenen informeert en wie kan besluiten tot aanpassing of stopzetting.

4.2 Wanneer wordt het gesprek gevoerd?

Het gesprek hoort niet pas plaats te vinden wanneer een model al is ontwikkeld en beleid feitelijk al is besloten. De toegevoegde waarde van het instrument ligt juist in het moment waarop binnen beleidsvorming expliciet wordt gekozen om modelgebruik een rol te geven. Dan kan nog worden besproken of modellering passend is, welke rol het model krijgt, welke aanvullende kennis nodig is en welke waarborgen vooraf moeten worden georganiseerd.

Tegelijkertijd is modelgebruik geen eenmalige keuze. Het gesprek kan nodig zijn bij de opdrachtverlening voor modelontwikkeling, bij de keuze om een bestaand model te hergebruiken, bij beleidsvorming waarin modeluitkomsten een rol krijgen, bij implementatie in uitvoering, bij publieke communicatie, bij signalen van schade of

controverse, en bij periodieke herbeoordeling. Het instrument moet daarom worden ingebed in zowel de beleidslevenscyclus als de modellevenscyclus.

De waarde van het instrument ligt uiteraard in de antwoorden die het oplevert, maar juist ook in een constatering dat vragen moeilijk of niet eenduidig kunnen worden beantwoord. Dat is niet per se een tekortkoming van het model of van de betrokkenen die het gesprek voeren. Modellen – zeker wanneer het gaat om complexe, samengestelde sociaal-technische systemen – kennen onvermijdelijk onzekerheden, afhankelijkheden en impliciete keuzes. Het is daarom juist te verwachten dat niet alle vragen eenvoudig te beantwoorden zijn. De vragen die (nog) niet beantwoord kunnen worden, zijn daarmee de meest waardevolle uitkomst van het gesprek. Zij maken zichtbaar waar de legitimiteit van het modelgebruik versterking behoeft, bijvoorbeeld door nadere onderbouwing, explicitering van aannames, aanvullende validatie, een scherpere afbakening van de bestuurlijke rol of duidelijkere afspraken over verantwoordelijkheid en herbeoordeling.

4.3 Institutionele inbedding

Het instrument moet voldoende institutioneel ingebed worden om daadwerkelijk te worden gebruikt, zonder te verworden tot een verplicht formulier dat vooral risico's juridisch afdekt. Afdwingbaarheid kan bijdragen aan navolgbaarheid, maar afdwingbaarheid op zichzelf brengt het noodzakelijke gesprek niet tot stand. Sterker nog: wanneer de prikkel vooral is om aan een verplicht format te voldoen, kan het instrument het gesprek juist verdringen.

Gebaseerd op wetenschappelijke literatuur,[10] schetsen we drie manieren waarop institutionele inbedding van het gespreksinstrument mogelijk is, elk met verschillende voor- en nadelen. Een minimumvariant is dat het instrument wordt aangeboden als vaste gespreksleidraad bij beleidsvoorstellen waarin modelgebruik een substantiële rol speelt, bijvoorbeeld als optionele bijlage bij het *Beleidskompas*. Dit creëert een expliciet moment waarop modelgebruik, rolclaim en grenzen worden besproken. Een versterkte variant is dat publieke organisaties het instrument koppelen aan bestaande besluitvormings-, kwaliteits- en verantwoordingsprocessen, zodat het gesprek verplicht wordt bij nieuw modelgebruik, hergebruik, doelverschuiving of hoge maatschappelijke impact. Een stelselmatige variant is dat overheidsbreed wordt geleerd van dossiers, bijsluiters, incidenten en goede praktijken, bijvoorbeeld via een kennisnetwerk of coördinerende functie of instituut voor verantwoord modelgebruik.

[10] Xue & Pang, 2022; Stahl, 2023; Van de Poel, 2020

Welke variant passend is, is een bestuurlijke keuze. Belangrijk is dat het instrument niet als geïsoleerd product wordt beheerd. Omdat het juist een boundary object is, bestaat anders het risico dat niemand eigenaar is, dat elementen selectief worden gebruikt of dat de bijsluiter losraakt van het dossier. Beheer, onderhoud, scholing en periodieke evaluatie zijn daarom onderdeel van de implementatieopgave.

5. Conclusie

Dit rapport vertrok vanuit de vraag naar een wetenschappelijke standaard voor het gebruik van modellen en algoritmes door de overheid. De analyse laat zien dat die vraag, alhoewel begrijpelijk, niet past op de onderliggende complexiteit en uitdagingen die gepaard gaan met het gebruik van modellen door de overheid. Modellen vervullen uiteenlopende rollen binnen beleid en uitvoering. Zij verschillen in doel, technische vorm, onzekerheid, context en maatschappelijke doorwerking. Daardoor verschillen ook de vormen van kwaliteit, assurance, transparantie, verantwoordelijkheid en tegenspraak die nodig zijn.

Onze centrale bevinding is dat de uitdaging niet primair ligt in het ontbreken van één technische norm, maar in het ontbreken van een gedeelde redenerings- en verantwoordingsstructuur voor modelgebruik in beleid. Verantwoord modelgebruik ontstaat niet in het model, maar in de wisselwerking tussen model, beleidsopgave, interpretatie, organisatie, publieke waarden en besluitvorming. Daarom verschuift de opgave van standaardisering naar bestuurlijke navolgbaarheid. De relevante vraag is niet of een model ‘goed’ of ‘fout’ is, maar of het gebruik van modelmatig geproduceerde kennis in een specifieke beleidscontext legitiem kan worden gedragen. Die legitimiteit vraagt een toereikende kennisbasis, een expliciete rolclaim, een duidelijke vertrouwensgrond, concrete waarborgen, passende informatie voor betrokkenen en een moment waarop het gebruik opnieuw kan worden beoordeeld of beëindigd.

Het voorgestelde gespreksinstrument biedt een praktisch aangrijpingspunt om deze elementen van expliciet te maken en te bespreken, en daarmee bij te dragen aan de verantwoorde en legitieme inzet van modellen in beleid. Het helpt de juiste mensen op het juiste moment de juiste vragen te stellen. Het vervangt bestaande standaarden, impact assessments, registers, audits of wetenschappelijke kwaliteitspraktijken niet, maar verbindt de inzichten daaruit in een bestuurlijke redenering over passend gebruik. Daarmee kan het instrument bijdragen aan modelgebruik dat openbaar, navolgbaar, verantwoordbaar en betwistbaar is.

De overheid kan en hoeft modelonzekerheid niet weg te standaardiseren. Zij moet onzekerheid, aannames, grenzen en normatieve keuzes zichtbaar maken en bestuurlijk hanteerbaar organiseren. Juist daarin ligt de meerwaarde van een gedeelde taal en een gestructureerd gesprek over verantwoord modelgebruik.

Bijlage 1. Gespreksinstrument voor verantwoord modelgebruik

Dit instrument ondersteunt beleidsmakers, opdrachtgevers, modelontwikkelaars, modelgebruikers, bestuurders, uitvoerders, toezichthouders en volksvertegenwoordigers bij het voeren en eventueel vastleggen van het gesprek over de inzet, betekenis en begrenzing van modellen.

Stap 0 - Wie moeten aan tafel zitten?

- Welke partijen zijn betrokken bij ontwikkeling, beheer, toepassing, interpretatie en besluitvorming?
- Welke juridische, beleidsmatige, technische, uitvoerings- en domeinexpertise is nodig?
- Welke groepen worden door het modelgebruik geraakt? Is betrokkenheid van geraakte groepen, maatschappelijke organisaties, toezicht of onafhankelijke expertise nodig?
- Wie is eigenaar van het gesprek en zorgt dat uitkomsten worden vastgelegd en opgevolgd?

Stap 1 - Is modellering een passende manier om het beleidsvraagstuk te ondersteunen?

- Wat is de beleidsopgave of het maatschappelijke vraagstuk waarvoor modelgebruik wordt overwogen?
- Welke kennisbehoefte, uitvoeringsopgave of besluitvraag ligt hieraan ten grondslag?
- Waarom is modellering passender dan alternatieven zoals aanvullende metingen, kwalitatieve kennis, participatie, experimenten, praktijkervaring of expliciete politieke keuzevorming?
- Welke publieke waarden, grondrechten, belangen of verdelingseffecten kunnen worden geraakt?
- Is er een reden om modelgebruik niet toe te passen, of alleen als scenario- of ondersteunend instrument?

Stap 2 - Op welke kennisbasis zijn de uitkomsten gebaseerd?

- Gaat het om een nieuw model, bestaand model, hergebruikte output, algoritmisch systeem, modelketen of combinatie van modellen en gegevensbronnen?
- Welke data, bronnen, definities, aannames, rekenregels en deskundigenoordelen liggen aan de uitkomsten ten grondslag?
- Welke aannames over gedrag, causaliteit, representativiteit, schaal, tijd of beleidsreacties zijn doorslaggevend?
- Welke kwaliteitsborging, validatie, verificatie, onzekerheidsanalyse, peer review, audit of standaard is toegepast?
- Welke organisaties, teams, leveranciers of kennisinstellingen leveren een bijdrage, en welke afhankelijkheden bestaan er?
- Waarop rust de vertrouwensgrond: technische kwaliteit, wetenschappelijke praktijk, organisatie, procedure, toezicht, participatie of combinatie daarvan?
- Zijn relevante juridische kaders, gedragscodes, AI- of algoritmestandaarden, DPIA/IAMA of domeinspecifieke normen betrokken?
-

Stap 3 - Welke betekenis kan aan de uitkomsten worden toegekend?

- Welke rol vervult het model? Is het bedoeld om te begrijpen, te verdelen, te anticiperen of te sturen?
- Voor welke situaties, schaalniveaus, doelgroepen en besluiten is het model ontwikkeld, en past dat bij het actuele gebruik?
- Welke conclusies kunnen verantwoord worden getrokken, en welke nadrukkelijk niet?
- Welke onzekerheden zijn gekwantificeerd, welke niet, en hoe worden niet-kwantificeerbare aspecten meegewogen?
- Welke aanvullende kennis, ervaring, juridische beoordeling, praktijkinformatie of maatschappelijke signalen zijn nodig naast de modeluitkomst?
- Welke normatieve of politieke keuzes mogen niet ongemerkt aan het model worden overgelaten?
- Wat is het verwerpingscriterium: wanneer mag de uitkomst deze rol niet meer vervullen?

Stap 4 - Hoe zijn verantwoordelijkheid, menselijke tussenkomst en tegenspraak georganiseerd?

- Wie is verantwoordelijk voor opdracht, ontwikkeling, beheer, toepassing, interpretatie, besluitvorming en gevolgen van modelgebruik?
- Welke menselijke oordeelsruimte bestaat er, en is die tijdig, geïnformeerd en bevoegd genoeg?
- Hoe worden modeluitkomsten beoordeeld op plausibiliteit, bruikbaarheid, ongelijke effecten en mogelijke schade?
- Welke ruimte bestaat er om af te wijken van modeluitkomsten, en hoe wordt afwijking vastgelegd?
- Welke escalatie-, correctie-, fallback- en stoproutes bestaan wanneer het model niet past, faalt of ongewenste gevolgen heeft?
- Hoe kunnen professionals, bestuurders, toezichthouders, burgers of maatschappelijke actoren aannames, uitkomsten of gebruik betwisten?
- Wie heeft de bevoegdheid om gebruik te beperken, te pauzeren of te beëindigen?

Stap 5 - Welke informatie is nodig om het gebruik te begrijpen en te beoordelen?

- Welke doelgroepen moeten modelgebruik kunnen begrijpen: beleidsmakers, bestuurders, Kamer, toezicht, uitvoering, burgers, bedrijven of maatschappelijke organisaties?
- Welke informatie hebben zij nodig over doel, rolclaim, data, aannames, onzekerheden, beperkingen, menselijke tussenkomst, verantwoordelijkheden en gevolgen?
- Welke informatie hoort in het interne governance-dossier en welke in een publieke bijsluiter, register of gerichte kennisgeving?
- Hoe wordt voorkomen dat openbaarmaking te technisch, te algemeen of misleidend geruststellend wordt?
- Hoe kan achteraf worden gereconstrueerd welke rol modeluitkomsten hebben gespeeld bij beleid of uitvoering?
- Welke mogelijkheden bestaan voor vragen, correctie, klacht, bezwaar, toezicht of parlementaire controle?

Stap 6 - Wanneer beoordelen we opnieuw of het model nog passend is?

- Welke veranderingen in beleid, wetgeving, data, techniek, modelversie, maatschappelijke context of gebruik kunnen herbeoordeling noodzakelijk maken?
- Wie signaleert veranderingen, incidenten, afwijkingen, maatschappelijke controverse of juridische uitspraken?
- Op welke momenten of met welke frequentie vindt herbeoordeling plaats?
- Wanneer openen veranderingen eerdere stappen opnieuw: beleidsopgave, kennisbasis, rolclaim, waarborgen, informatie of stoplogica?
- Wanneer wordt modelgebruik aangepast, beperkt, aangevuld, gepauzeerd of beëindigd?
- Wanneer is een nieuwe of geactualiseerde DPIA, IAMA, audit, externe toets of publieke toelichting nodig?
- Hoe blijven dossier en bijsluiter actueel?

Bijlage 2. Totstandkoming, scope en verhouding tot de motie

Dit rapport is opgesteld naar aanleiding van de politieke vraag naar een wetenschappelijke standaard voor het gebruik van modellen en algoritmes door de overheid. De motie van de leden Omtzigt en Six Dijkstra vraagt hoe zo'n standaard wordt samengesteld, hoe afdwingbaar wordt gemaakt dat modellen en algoritmes openbaar en navolgbaar zijn, en hoe een bijsluiter voor modellen zal worden opgesteld. Met die vraag is het onderzoek begonnen. De uiteindelijke focus van het rapport is echter verschoven: niet omdat de vraag naar standaardisering onbelangrijk is, maar omdat uit het onderzoek blijkt dat één uniforme wetenschappelijke standaard de kern van het probleem niet goed raakt.

Inhoudelijk is het onderzoek breed opgezet. Wij organiseerden een multidisciplinair panel op de Data for Policy conferentie in Den Haag op 12 en 13 juni 2025. Aan dit panel namen wetenschappers deel, evenals een vertegenwoordiger van een onafhankelijk kennisinstituut met expertise op duurzaamheidsmodellen en een expert die betrokken is bij standaardisatie rondom algoritmische modellen. Daarnaast voerden wij meerdere gesprekken met personen verbonden aan de NEN-normcommissie AI. Wij namen deel aan panel 5, "Establishing robust AI Standards", tijdens de International AI Summit in Brussel en spraken daar met experts op het gebied van CEN/CENELEC en ISO/IEC over het ontwikkelen van standaarden voor AI en algoritmische systemen.

In totaal zijn, naast bovenstaande activiteiten, tien expertinterviews en -gesprekken gevoerd. Deze gesprekken vonden plaats met wetenschappers op het gebied van complexe modellen voor beleid, projectleiders rondom modeltransparantie, actieonderzoekers op het gebied van AI-governance en standaarden, standaardisatie-experts en personen vanuit de overheid, waaronder vanuit toezichthoudend perspectief. Daarnaast zijn twee bijeenkomsten met een adviesgroep georganiseerd. Deze bijeenkomsten dienden als groepsinterview en als klankbord. Bij de eerste bijeenkomst waren zeven experts aanwezig, bij de tweede zes. De deelnemers waren werkzaam bij de overheid of bij onafhankelijke expertinstituten rondom complexe modellen of toezicht daarop; daarnaast was een onafhankelijk lid betrokken.

Naast deze gesprekken is literatuuronderzoek uitgevoerd en is gekeken naar praktijken in andere landen. Daarbij is onder meer gekeken naar wetenschappelijke literatuur over modelvalidatie en -transparantie, onzekerheid en onzekerheidsanalyse, post-normal

science, participatieve modellering, boundary objects, transparantie, algoritmische accountability, modeldocumentatie, compound modelling, modelketens, interoperabiliteit en governance van AI-systemen. Ook zijn bestaande instrumenten en standaarden bekeken, zoals impact assessments, algoritmeregisters, model cards, datasheets, internationale AI-raamwerken en opkomende standaardisatieprocessen. Al die onderwerpen verdienen afzonderlijk aandacht, maar dat is buiten de scope van dit rapport. In dit rapport zijn die thema's verkend en waar relevant gebruikt, maar niet elk thema is uitgewerkt tot afzonderlijke lijn.

De oorspronkelijke vraag naar een afdwingbare wetenschappelijke standaard heeft in dit traject een centrale rol gespeeld. Juist door deze vraag serieus te nemen, werd zichtbaar waar de grenzen ervan liggen. Voor technische onderdelen van modelontwikkeling kunnen standaarden, normen en good practices waardevol zijn. Engineers, modelleurs en domeinexperts kunnen afspraken maken over documentatie, validatie, onzekerheidsanalyse, interoperabiliteit, biasdetectie, reproduceerbaarheid of kwaliteitsborging. In de wereld van AI en algoritmen ontstaan bovendien standaarden en wettelijke vereisten die, afhankelijk van context en systeemtype, relevant kunnen zijn voor publieke organisaties.

Tegelijkertijd werd in gesprekken en literatuur duidelijk dat het meest gevoelige deel van publiek modelgebruik niet volledig technisch is. Modellen en algoritmen in beleid zijn sociotechnische instrumenten. Zij combineren technische keuzes met institutionele context, juridische normen, politieke afwegingen, publieke waarden en maatschappelijke gevolgen. De vraag wat aanvaardbare schade is, welke verdeling rechtvaardig is, welk risico proportioneel is of welke belangen zwaarder wegen, kan niet door een wetenschappelijke standaard worden beslecht. Daarvoor bestaan wel wetenschappelijke inzichten over goede processen, transparantie, participatie en verantwoording, maar geen inhoudelijke standaard die de politieke en normatieve beoordeling overneemt.

Klassieke (technische) standaardisatie werkt vaak via consensus over technische specificaties. Echter, onze verkenning van standaardisatie van algoritmische modellen laat zien dat vragen over maatschappelijke schade, discriminatie, verantwoordelijkheid, rechten en democratische controle, centraal staan. Daarmee komen standaardisatieorganen nadrukkelijker in aanraking met sociotechnische vraagstukken, waartoe proces en aanpak zich niet altijd even goed lenen. Voor het technische deel kan standaardisatie veel

betekenen; voor het normatieve en politieke deel is zij slechts een deel van het governanceproces.

De gekozen focus is daarom een eerste, noodzakelijke stap: het organiseren van het gesprek tussen de beleidskant en de modelkant. Goed gebruik van modellen is niet alleen een opdracht aan modelmakers en modeleigenaren; het is minstens zozeer een opdracht aan beleidsmakers, bestuurders en uitvoerders. Zij moeten expliciet maken welke rol modeluitkomsten krijgen, welke kennisclaim zij daaraan verbinden, welke normatieve keuzes zij zelf moeten maken, welke aanvullende kennis nodig is en wanneer het gebruik opnieuw moet worden beoordeeld. Als dat gesprek goed wordt gevoerd, volgt daaruit ook welke technische beoordeling, standaard, audit, bijsluiter, registratie of juridische toets in een concrete casus nodig is.

In verhouding tot de motie betekent dit dat het rapport niet uitkomt op één afdwingbare wetenschappelijke standaard, maar op een voorstel voor het articuleren van de verantwoordingsroute. Zelfs als iets dergelijks bestuurlijk of institutioneel verplicht worden gesteld bij modelgebruik met voldoende impact, legitimeert het modelgebruik niet automatisch. Het maakt zichtbaar welke vragen moeten worden beantwoord en welke informatie beschikbaar moet zijn om modelgebruik openbaar, navolgbaar, verantwoordbaar en betwistbaar te maken.

Referenties en geraadpleegde bronnen

Adviescollege Meten en Berekenen Stikstof (2020). *Meer meten, robuuster rekenen*.

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

Alon-Barkat, S., & Busuioc, M. (2023). Human–AI interactions in public sector decision making: "Automation bias" and "selective adherence" to algorithmic advice. Journal of Public Administration Research and Theory, 33(1), 153–169. https://doi.org/10.1093/jopart/muac007

Alon-Barkat, S., Busuioc, M., Schwoerer, K., & Weißmüller, K. S. (2025). Algorithmic discrimination in public service provision: Understanding citizens' attribution of responsibility for human versus algorithmic discriminatory outcomes. Journal of Public Administration Research and Theory, 35(4), 469–488. https://doi.org/10.1093/jopart/muaf024

Andersson, J. D. (2023). Ecologies of integrated modeling: Configuring policy-relevance in Swedish climate governance. Frontiers in Climate, 5, Article 1159860. https://doi.org/10.3389/fclim.2023.1159860

Autoriteit Persoonsgegevens. (2024). Onderzoeksrapport fraudeaanpak DUO. Autoriteit Persoonsgegevens. https://www.autoriteitpersoonsgegevens.nl/documenten/onderzoeksrapport-fraudeaanpak-duo

Bell, A. R., Troost, C., Filatova, T., Vincenot, C. E., Qin, C.-Z., Fiedler, S., Lemmen, C., Lynch, I., Pierce, M., Edmonds, B., McIntire, E., Wang, H.-H., Koralewski, T. E., Grant, W. E., Wang, M., Black, B., & Grimm, V. (2026). The PAVE pathway: How to talk about models with people who built them, and people who didn't. BioScience. Advance online publication. https://doi.org/10.1093/biosci/biag010

Bignami, F. (2022). Artificial intelligence accountability of public administration. The American Journal of Comparative Law, 70(Supplement_1), i312–i346. https://doi.org/10.1093/ajcl/avac012

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

Brugnach, M., Tagg, A., Keil, F., & de Lange, W. J. (2007). Uncertainty matters: computer models at the science-policy interface. Water Resources Management, 21(7), 1075-1090. https://doi.org/10.1007/s11269-006-9099-y

Busuioc, M. (2021). Accountable artificial intelligence: Holding algorithms to account. Public Administration Review, 81(5), 825–836. https://doi.org/10.1111/puar.13293

Cobbe, J., Veale, M., & Singh, J. (2023). Understanding accountability in algorithmic supply chains. Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency.
pp. 1186 – 119.7. https://doi.org/10.1145/3593013.359407

Council of Europe. (2026). HUDERIA methodology and model: Context-based risk analysis (COBRA) resources. Council of Europe. https://book.coe.int/en/artificial-intelligence/12520-pdf-huderia-methodology-and-model.html

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning [Preprint]. arXiv. https://doi.org/10.48550/arXiv.1702.08608

Duda, S., Hofmann, P., Urbach, N., Völter, F., & Zwickel, A. (2024). The impact of resource allocation on the machine learning lifecycle: Bridging the gap between software engineering and management. Business & Information Systems Engineering, 66(2), 203–219. https://doi.org/10.1007/s12599-023-00840-4

ECNL & DIHR. (2025). A guide to fundamental rights impact assessments ("FRIA") under the EU Artificial Intelligence Act. European Center for Not-for-Profit Law & Danish Institute for Human Rights.

Edmonds, B. (2017). Different modelling purposes. In B. Edmonds & R. Meyer (Eds.), Simulating social complexity: A handbook. Springer.

Edmonds, B., Le Page, C., Bithell, M., Chattoe-Brown, E., Grimm, V., Meyer, R., Montañola-Sales, C., Ormerod, P., Root, H., & Squazzoni, F. (2019). Different modelling purposes. Journal of Artificial Societies and Social Simulation, 22(3), Article 6. https://doi.org/10.18564/jasss.3993

Eubanks, V. (2018). Automating inequality: How high-tech tools profile, police, and punish the poor. St. Martin's Press.

Filatova, T., Akkerman, J., Bosello, F., Chatzivasileiadis, T., Cortés Arbués, I., Ghorbani, A., Ivanova, O., Knittel, N., Kwakkel, J., Lamperti, F., Magliocca, N. R., Marangoni, G., Nabernegg, S., Pichler, A., Poujon, A., Safarzynska, K., Taberna, A., van Sluisveld, M. A. E., Verbeek, L., & Wei, T. (2025). The power of bridging decision scales: Model coupling for advanced climate policy analysis. Proceedings of the National Academy of Sciences, 122(38), Article e2411592122. https://doi.org/10.1073/pnas.2411592122

Funtowicz, S. O., & Ravetz, J. R. (1993). Science for the Post-Normal Age. *FUTURES*, 739-755.

Funtowicz, S. O., & Ravetz, J. R. (1994). Uncertainty, complexity and post-normal science. *Environmental Toxicology and Chemistry*, *13*(12), 1881-1885. https://doi.org/10.1002/etc.5620131203

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

Gerard, C., & Beunen, R. (2025). Begrenzingen aan AERIUS: Kritische reactie op het artikel “Welke begrenzingen zijn er aan het toepassen van het stikstofrekenmodel AERIUS?”. Tijdschrift Lucht, 21(2), 14–19.

Giest, S., & Klievink, B. (2024). More than a digital system: How AI is changing the role of bureaucrats in different organizational contexts. *Public Management Review*, *26*(2), 1-20. https://doi.org/10.1080/14719037.2022.2095001

Government Digital Service. (2025). Algorithmic transparency recording standard: Guidance for public sector bodies. GOV.UK. https://www.gov.uk/government/publications/guidance-for-organisations-using-the-algorithmic-transparency-recording-standard/algorithmic-transparency-recording-standard-guidance-for-public-sector-bodies

Green, B., & Chen, Y. (2019). The principles and limits of algorithm-in-the-loop decision making. Proceedings of the ACM on Human-Computer Interaction, 3(CSCW), Article 50. https://doi.org/10.1145/3359152

Grimm, V., Augusiak, J., Focks, A., Frank, B. M., Gabsi, F., Johnston, A. S. A., Liu, C., Martin, B. T., Meli, M., Radchuk, V., Thorbek, P., & Railsback, S. F. (2014). Towards better modelling and decision support: Documenting model development, testing, and analysis using TRACE. Ecological Modelling, 280, 129–139. https://doi.org/10.1016/j.ecolmodel.2014.01.018

Grimm, V., Railsback, S. F., Vincenot, C. E., Berger, U., Gallagher, C., DeAngelis, D. L., Edmonds, B., Ge, J., Giske, J., Groeneveld, J., Johnston, A. S. A., Milles, A., Nabe-Nielsen, J., Polhill, J. G., Radchuk, V., Rohwäder, M.-S., Stillman, R. A., Thiele, J. C., & Ayllón, D. (2020). The ODD protocol for describing agent-based and other simulation models: A second update to improve clarity, replication, and structural realism. Journal of Artificial Societies and Social Simulation, 23(2), Article 7. https://doi.org/10.18564/jasss.4259

Grimmelikhuijsen, S., & Meijer, A. (2022). Legitimacy of algorithmic decision-making: Six threats and the need for a calibrated institutional response. Perspectives on Public Management and Governance, 5(3), 232–242. https://doi.org/10.1093/ppmgov/gvac008

Grothe, M., & Koomen, E. (2025). Verkenning interoperabiliteit van rekenmodellen en digitale tweelingen. Geonovum.

Hall, D. M., Lazarus, E. D., & Swannack, T. M. (2014). Strategies for communicating systems models. Environmental Modelling & Software, 55, 70-76. https://doi.org/10.1016/j.envsoft.2014.01.007

Infocomm Media Development Authority & AI Verify Foundation. (2025). AI Verify testing framework: For traditional and generative AI. AI Verify Foundation.

Informatiepunt Leefomgeving. (z.d.). Nationaal Water Model. Geraadpleegd op 18 mei 2026, van https://iplo.nl/thema/water/applicaties-modellen/watermanagementmodellen/nationaal-water-model/

International Organization for Standardization & International Electrotechnical Commission. (2023). ISO/IEC 42001:2023: Information technology—Artificial intelligence—Management system. ISO.

Kemper, J., & Kolkman, D. (2019). Transparent to whom? No algorithmic accountability without a critical audience. *Information, Communication & Society*, *22*(14), 2081–2096. https://doi.org/10.1080/1369118X.2018.1477967

Kloprogge, P., Van Der Sluijs, J. P., & Petersen, A. C. (2011). A method for the analysis of assumptions in model-based environmental assessments. *Environmental Modelling & Software*, *26*(3), 289-301. https://doi.org/10.1016/j.envsoft.2009.06.009

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.

Lorenz, L. (2023). The hybrid work of public sector data scientists. Journal of Professions and Organization, 10(3), 226–242. https://doi.org/10.1093/jpo/joad017

Meester, R. (2025). *De illusie van een betrouwbare stikstof-modelwerkelijkheid*

Meijer, A., Grimmelikhuijsen, S., & Bovens, M. (2021). De legitimiteit van het algoritmisch bestuur: Een systematisch overzicht van bedreigingen en oplossingsrichtingen. Nederlands Juristenblad, 18, 1470–1478.

Meijer, A., Lorenz, L., & Wessels, M. (2021). Algorithmization of bureaucratic organizations: Using a practice lens to study how context shapes predictive policing systems. Public Administration Review, 81(5), 837–846. https://doi.org/10.1111/puar.13391

Mergel, I., Dickinson, H., Stenvall, J., & Gascó, M. (2023). Implementing AI in the public sector. Public Management Review. Advance online publication. https://doi.org/10.1080/14719037.2023.2231950

Ministerie van Infrastructuur en Waterstaat. (2019). Nationaal Water Model: Een gezamenlijk model voor waterbeleid in Nederland. Ministerie van Infrastructuur en Waterstaat. https://iplo.nl/publish/pages/132724/2020_02_11_brochure_nationaal_watermodel_web_1.pdf

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. In Proceedings of the Conference on Fairness, Accountability, and Transparency (pp. 220–229). Association for Computing Machinery. https://doi.org/10.1145/3287560.3287596

Mougan, C., Morlock, L., Aguirre, J., Black, J. R. M., Brauner, J., Campos, S., Dev, S., Fernández Llorca, D., Franzin, A., Fritz, M., Gómez, E., Grosse-Holz, F., Hamilton, E., Hasin, M., Hernandez-Orallo, J., Lahav, D., Massarelli, L., Mavroudis, V., Murray, M., ... Schellaert, W. (2026). The science and practice of proportionality in AI risk evaluations. Science, 391(6787), 769–771. https://doi.org/10.1126/science.aea3835

Natarajan, S., Mathur, S., Sidheekh, S., Stammer, W., & Kersting, K. (2025). Human-in-the-loop or AI-in-the-loop? Automate or Collaborate?. *Proceedings of the AAAI Conference on Artificial Intelligence*, *39*(27), 28594–28600. https://doi.org/10.1609/aaai.v39i27.35083

National Institute of Standards and Technology. (2023). Artificial intelligence risk management framework (AI RMF 1.0) (NIST AI 100-1). U.S. Department of Commerce. https://doi.org/10.6028/NIST.AI.100-1

Nikiforova, A., Lnenicka, M., Melin, U., Valle-Cruz, D., Gill, A., Casiano Flores, C., Sirait, E., Luterek, M., Dreyling, R. M., & Tesarova, B. (2025). Responsible AI adoption in the public sector: A data-centric taxonomy of AI adoption challenges [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2510.09634

Organisation for Economic Co-operation and Development. (2022). OECD framework for the classification of AI systems. OECD Publishing. https://doi.org/10.1787/cb6d9eca-en

Oreskes, N., Shrader-Frechette, K., & Belitz, K. (1994). Verification, validation, and confirmation of numerical models in the earth sciences. Science, 263(5147), 641–646. https://doi.org/10.1126/science.263.5147.641

Oswald, M., Chambers, L., Goodman, E. P., Ugwudike, P., & Zilka, M. (2022). The UK algorithmic transparency standard: A qualitative analysis of police perspectives. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4155549

Parker, W. S. (2018). The significance of robust climate projections. In E. A. Lloyd & E. Winsberg (Eds.), Climate modelling: Philosophical and conceptual issues (pp. 273–296). Springer.

Pullen, E., Ruijer, E., & Meijer, A. (2026). Governing ethics for the digital transformation: Developing, testing, and validating a framework. Government Information Quarterly, 43(2), Article 102117. https://doi.org/10.1016/j.giq.2026.102117

Raad van State. (2014, 15 mei). Ontwerpbesluit houdende regels voor fraudeaanpak door gegevensuitwisselingen en het effectief gebruik van binnen de overheid bekend zijnde gegevens (Besluit SyRI), met nota van toelichting (Advies W12.14.0102/III). Raad van State. https://www.raadvanstate.nl/adviezen/@63143/w12-14-0102-iii/

Rechtbank Den Haag. (2020, 5 februari). ECLI:NL:RBDHA:2020:865. https://uitspraken.rechtspraak.nl/#!/details?id=ECLI:NL:RBDHA:2020:865

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), OJ L, 2024/1689. (2024).

Rijkswaterstaat. (z.d.). Watermodellen. Geraadpleegd op 18 mei 2026, van https://www.rijkswaterstaat.nl/zakelijk/open-data/modellen-en-applicaties/watermodellen

Robertson, S. (2021). Transparency, trust, and integrated assessment models: An ethical consideration for the Intergovernmental Panel on Climate Change. WIREs Climate Change, 12, Article e679. https://doi.org/10.1002/wcc.679

Rodés-Bachs, C., Sampedro, J., Frilingou, N., Gardumi, F., & Lo Giudice, C. (2025). Open science practices in integrated assessment models. Open Research Europe, 5, Article 12. https://doi.org/10.12688/openreseurope.18824.2

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

Schmitz, C., Rystrøm, J., & Batzner, J. (2025). Oversight structures for agentic AI in public-sector organizations [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2506.04836

Schröter, K., Schweizer, P.-J., Gräler, B., Cumiskey, L., Bharwani, S., Parviainen, J., Kropf, C. M., Håkansson, V. W., Drews, M., Irvine, T., Dondi, C., Apel, H., Löhrlein, J., Hochrainer-Stigler, S., Bagli, S., Huszti, L., Genillard, C., Unguendoli, S., Hattermann, F., & Steinhausen, M. (2025). Invited perspectives: Fostering interoperability of data, models, communication, and governance for disaster resilience through transdisciplinary knowledge co-production. Natural Hazards and Earth System Sciences, 25, 3055–3073. https://doi.org/10.5194/nhess-25-3055-2025

Selbst, A. D., boyd, d., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. In Proceedings of the 2019 Conference on Fairness, Accountability, and Transparency (pp. 59–68). Association for Computing Machinery. https://doi.org/10.1145/3287560.3287598

Stahl, B. C. (2023). Embedding responsibility in intelligent systems: from AI ethics to responsible AI ecosystems. Scientific Reports, 13(1), 7586. https://doi.org/10.1038/s41598-023-34622-w

Star, S. L. (2010). This is Not a Boundary Object: Reflections on the Origin of a Concept. *Science, Technology, & Human Values*, *35*(5), 601–617. http://www.jstor.org/stable/25746386

Stone, D. (2012). Policy paradox: The art of political decision making (3rd ed.). W.W. Norton & Company

Taeihagh, A. (2025). Governance of generative AI. Policy and Society, 44(1), 1–22. https://doi.org/10.1093/polsoc/puaf001

Treasury Board of Canada Secretariat. (2025). Directive on automated decision-making. Government of Canada. https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592

Trondal, J., Haslerud, G., & Kühn, N. S. (2021). The robustness of national agency governance in integrated administrative systems: Evidence from a large-scale study. *Public Administration Review, 81*(1), 121–136. https://doi.org/10.1111/puar.13309

Van de Poel, I. (2020). Embedding values in artificial intelligence (AI) systems. Minds and machines, 30(3), 385-409. https://doi.org/10.1007/s11023-020-09537-4

Vereniging van Nederlandse Gemeenten. (z.d.). Werk en inkomen digitaal (Suwinet). Geraadpleegd op 18 mei 2026, van https://vng.nl/projecten/werk-en-inkomen-digitaal-suwinet

van der Sluijs JP, Craye M, Funtowicz S, Kloprogge P, Ravetz J, Risbey J. Combining quantitative and qualitative measures of uncertainty in model-based environmental assessment: the NUSAP system. Risk Anal. 2005 Apr;25(2):481-92. doi: 10.1111/j.1539-6924.2005.00604.x. PMID: 15876219.

Veale, M., & Edwards, L. (2018). Clarity, surprises, and further questions in the Article 29 Working Party draft guidance on automated decision-making and profiling. Computer Law & Security Review, 34(2), 398–404. https://doi.org/10.1016/j.clsr.2017.12.002

Veale, M., Van Kleek, M., & Binns, R. (2018). Fairness and accountability design needs for algorithmic support in high-stakes public sector decision-making. In Proceedings of the 2018 CHI Conference on Human Factors in

Computing Systems (Article 440, pp. 1–14). Association for Computing Machinery. https://doi.org/10.1145/3173574.3174014

Voinov, A., Kolagani, N., McCall, M. K., Glynn, P. D., Kragt, M. E., Ostermann, F. O., Pierce, S. A., & Ramu, P. (2016). Modelling with stakeholders - next generation. Environmental Modelling & Software, 77, 196-220. https://doi.org/10.1016/j.envsoft.2015.11.016

Vydra, S., & Klievink, B. (2019). Techno-optimism and policy-pessimism in the public sector big data debate. Government Information Quarterly, 36(4), 101383. https://doi.org/10.1016/j.giq.2019.05.010

Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. International Data Privacy Law, 7(2), 76–99. https://doi.org/10.1093/idpl/ipx005

Wardekker, J. A., Van Der Sluijs, J. P., Janssen, P. H. M., Kloprogge, P., & Petersen, A. C. (2008). Uncertainty communication in environmental assessments: Views from the Dutch science-policy interface. *Environmental Science & Policy*, *11*(7), 627-641. https://doi.org/10.1016/j.envsci.2008.05.005

Widlak, A. C., & Peeters, R. (2025). A theory of the infrastructure-level bureaucracy: Understanding the consequences of data-exchange for procedural justice, organizational decision-making, and data itself. Government Information Quarterly, 42(2), Article 102021. https://doi.org/10.1016/j.giq.2025.102021

Xue, L., & Pang, Z. (2022). Ethical governance of artificial intelligence: An integrated analytical framework. Journal of Digital Economy, 1(1), 44-52. https://doi.org/10.1016/j.jdec.2022.08.003