WHITEPAPER · AI IN DENKENDE BEROEPEN

Waar rust deze conclusie op?

Een organisatie kan aan alle kwaliteitseisen voldoen en toch niet laten zien waarop een aanbeveling rust. Wie dat wel kan laten zien, mag ook sneller. Over die twee kanten gaat deze whitepaper.

VERSIE

6.3 · 20 augustus 2026

VOOR

de denkende beroepen: onderzoek, advies, uitgeverij en onderwijs, en hun opdrachtgevers

GRONDSLAG

elke bewering gelabeld, verantwoording achterin

INHOUD

IN HET KORT

Vroeg of laat doet een opdrachtgever navraag over een aanbeveling uit een rapport en wil weten waar die op gebaseerd is. Dat antwoord bestaat alleen wanneer tijdens het onderzoek is vastgelegd welke bevindingen de aanbeveling ondersteunen en op welk materiaal die bevindingen leunen. Geen enkel kader voor kennisorganisaties vraagt om traceerbaarheid. ISO 20252, de norm voor markt-, opinie- en sociaal onderzoek, waar beleidsonderzoek onder valt, toetst of het onderzoek als geheel zo is vastgelegd dat het reproduceerbaar is. Waar één conclusie vandaan komt, valt buiten die toets.

AI maakt het risico nog groter, want een verzonnen bron is met het blote oog niet van een echte te onderscheiden. Bij een transcript valt dat op te vangen: daar ligt een origineel naast. Bij een aanbeveling niet. Die ontstaat in het hoofd van de onderzoeker, en daar houdt het spoor op.

Een beter taalmodel lost dit niet op. Een andere inrichting van het proces wel. En die inrichting beschermt niet alleen: dezelfde afspraak die het risico wegneemt, maakt de winst vrij, want wat aantoonbaar op materiaal rust, mag sneller door. Drie van de maatregelen in deel III kosten geen geld en kunnen deze week ingaan. En soms is de uitkomst dat AI ergens niets toevoegt. Ook dat zetten wij op papier.

INLEIDING

De vraag die pas maanden later komt

Een kennisorganisatie legt veel vast. De methode staat beschreven, de transcripten zijn bewaard, het codeboek en de conceptversies staan in het archief. Elk onderdeel van de keten is er. Wat de werkwijze niet vanzelf oplevert, is het verband over die keten heen: op welke bevindingen een aanbeveling is gebaseerd, en op welk materiaal die bevindingen steunen.

Zolang niemand ernaar vraagt, valt dat niet op. De vraag komt vaak pas maanden na de oplevering, wanneer het rapport ergens begint te schuren: een gemeenteraad debatteert over één aanbeveling, en de opdrachtgever wil weten op welke interviews en welke analyse die steunt. Of dat antwoord er dan is, hangt niet af van de zorgvuldigheid van het werk, maar van één inrichtingskeuze: is het verband tijdens het onderzoek vastgelegd, of moet het achteraf uit het archief worden gereconstrueerd?

Deze whitepaper gaat over dat ontbrekende verband.

Eén afspraak vooraf. Elke feitelijke bewering in dit document draagt een label, BEWEZEN, AANNEMELIJK of OMSTREDEN, en de verantwoording per bron staat achterin. Zo hoeft de lezer dit document niet klakkeloos aan te nemen. Dat is de bedoeling.

Elk stuk is netjes gearchiveerd. Alleen het verband tussen de stukken ligt nergens vast.

Herleidbaar is iets anders dan reproduceerbaar

Twee begrippen lijken op elkaar en lopen in de praktijk door elkaar. Reproduceerbaar betekent dat een onderzoek met dezelfde methode kan worden herhaald en tot vergelijkbare uitkomsten leidt. Herleidbaar betekent dat elke conclusie in het rapport terug te voeren is naar het materiaal waarop die gebaseerd is. Kwaliteitssystemen borgen het eerste. Opdrachtgevers vragen naar het tweede.

Het overkomt de kantoren die het kunnen weten

Hoe dat misgaat, laat het vak zelf zien.

- **Deloitte Australië** betaalde een deel van een overheidsopdracht terug nadat in het rapport verzonnen bronnen en een verzonnen citaat uit een rechterlijke uitspraak waren aangetroffen. BEWEZEN
- **KPMG** haalde een internationaal adviesrapport over AI offline nadat onderzoekers van GPTZero de 45 bronverwijzingen hadden nagelopen: vijf klopten, de rest bleek deels verzonnen, verdraaid of te vaag om te controleren. BEWEZEN
- Een **Nederlandse advocaat** voerde bij de rechtbank jurisprudentie op die niet bestond, kreeg dat van de rechter te horen, en kwam in een volgende zaak bij dezelfde rechtbank

opnieuw met niet-bestaande uitspraken. AANNEMELIJK

Sinds dit voorjaar is het rijtje compleet: ook EY Canada (mei 2026) en PwC Middle East (eind juli 2026, vier rapporten) moesten werk terugtrekken of corrigeren nadat verzonnen bronnen waren aangetroffen. Daarmee had elk van de vier grote kantoren binnen een jaar een eigen geval. AANNEMELIJK

Geen van deze stukken viel op voordat iemand de bronnen ging nalopen. Daar zit de kern: het was werk dat eruitzag als af werk, gemaakt door mensen die hun vak verstaan.

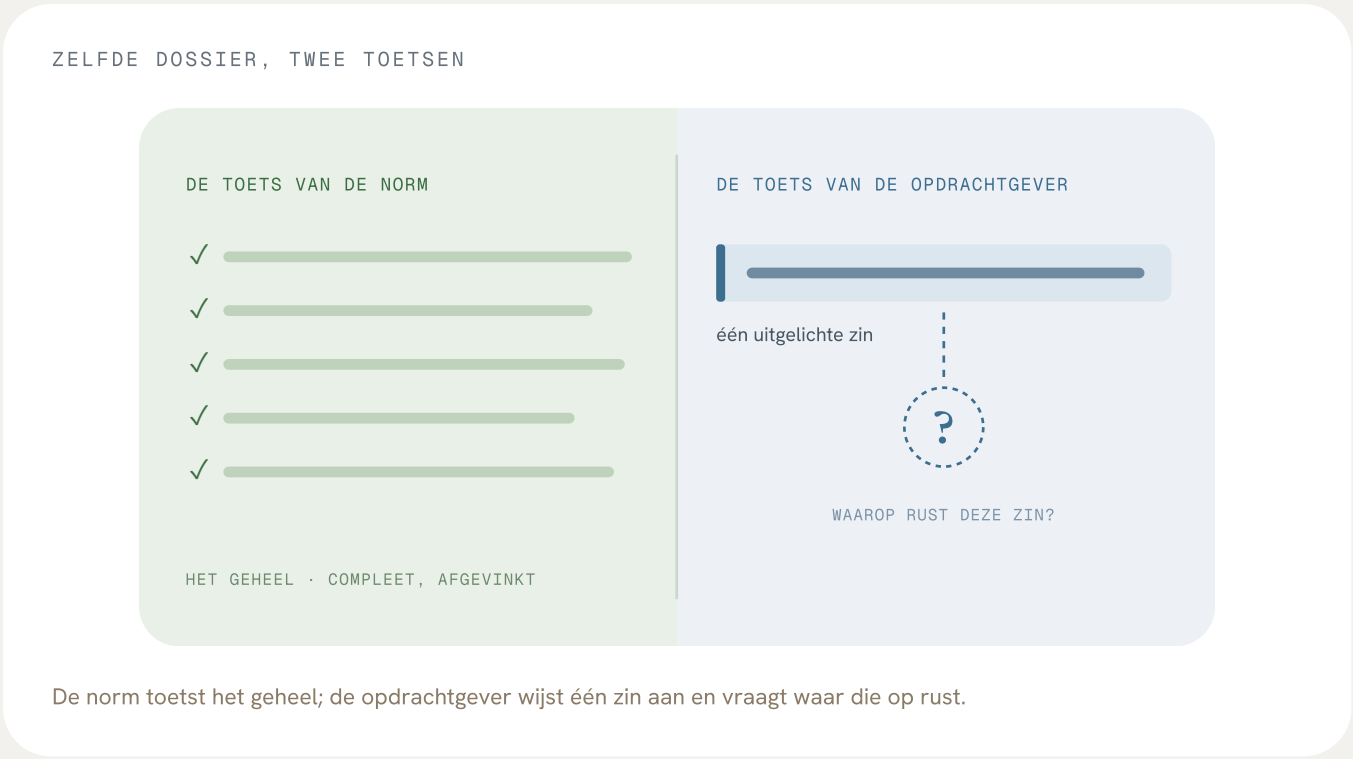
Wat de norm toetst en wat de opdrachtgever vraagt

Traceerbaarheid is in deze branche geen nieuw woord. Het staat in de kwaliteitshandboeken en in de eigen kwaliteitsvisie van vrijwel elke kennisorganisatie: conclusies moeten traceerbaar zijn, de onderzoeker blijft eindverantwoordelijk, AI ondersteunt de expertise en vervangt die niet. Dat is serieus bedoeld en het wordt serieus genomen.

Het misverstand zit in wat er vervolgens getoetst wordt. ISO 20252 vraagt of methode, uitvoering en analyse zo zijn vastgelegd dat het onderzoek herhaalbaar is. Dat is een toets op het geheel. De opdrachtgever toetst het deel: die pikt er één conclusie uit en wil weten waar die vandaan komt, en of er een taalmodel aan te pas is gekomen.

Een organisatie kan dus gecertificeerd zijn, integer werken en op die ene vraag toch het antwoord schuldig blijven. Dat gat bestond al voordat AI bestond. AI heeft het zichtbaar gemaakt en groter.

Hoe vaak dat voorkomt, heeft niemand gemeten, en een percentage zou hier dus verzonnen zijn. In plaats daarvan staat er een proef, en die kost een kwartier: pak het laatst opgeleverde rapport, kies één aanbeveling en probeer te laten zien op welke bevindingen die rust en op welk materiaal die bevindingen steunen. Lukt dat binnen een kwartier, dan is deze whitepaper een bevestiging van wat er al staat. Lukt het niet, dan staat hierna waar het spoor breekt en wat daaraan te doen is.



De norm toetst het geheel; de opdrachtgever wijst één zin aan en vraagt waar die op rust.

Hoe dit document verder is opgebouwd

Deel I beschrijft de stand van zaken: waar AI het onderzoeksproces al is binnengekomen, welke regels gelden en wat de bestaande kwaliteitskaders wel en niet afdekken.

Deel II laat zien waar het risico werkelijk zit, aan de hand van SEAL: de vier fasen die kenniswerk doorloopt en waarin de herleidbaarheid van een bewering verloren kan gaan. Het risico zit niet bij de transcriptie, en ook niet bij de bevinding die naast het eigen materiaal te leggen is, maar bij de stap van bevinding naar aanbeveling.

Deel III is praktisch: drie maatregelen die geen budget vragen, en TRACES, de zes lagen waarop je een systeem of een leverancier toetst. Soms is de uitkomst van die toets dat AI op een plek niets toevoegt. Ook dat is een uitkomst.

I

DEEL I

Wat er speelt

HOOFDSTUK 1

AI wordt al gebruikt in de analyse zelf, meestal buiten zicht

Vier momenten uit de dagelijkse praktijk op een onderzoeksafdeling.

- Een **onderzoeker** plakt een interviewtranscript in een privé-account van ChatGPT om het te laten opschonen.
- Een **projectleider** laat een vertrouwelijk beleidsstuk van de opdrachtgever samenvatten omdat de bespreking om twee uur begint.
- Een **junior** besteedt de literatuurscan uit aan een taalmodel en neemt de treffers over zonder ze na te trekken.
- Een **eindredacteur** laat het slothoofdstuk herschrijven, de avond voor de oplevering.

Geen van deze vier handelingen staat ergens geregistreerd. Ze verschijnen in geen projectdossier en geen kwaliteitsrapportage, en juist daardoor is de omvang van het gebruik zo lastig te meten.

Wie het met een enquête probeert, meet te laag. Een medewerker die met een privé-account werkt, meldt dat niet snel in een vragenlijst van de eigen werkgever. Gedragsdata komen dichter bij de werkelijkheid.

Twee metingen, onafhankelijk van elkaar, komen op hetzelfde uit.

- **LayerX**, telemetrie op wat er feitelijk in browsers gebeurt: ongeveer 45 procent van de werknemers gebruikt generatieve AI. Het grootste deel loopt via onbeheerde privé-accounts, en een aanzienlijk deel van de geüploade bestanden bevat gevoelige of persoonlijke gegevens. AANNEMELIJK
- **Verizon**, Data Breach Investigations Report 2026: op basis van ruim 850.000 gelogde gebeurtenissen hetzelfde percentage, drie keer zoveel als een jaar eerder. AANNEMELIJK

WAT DIT CIJFER WEL EN NIET ZEGT

De meting van LayerX gaat over werknemers in het algemeen, niet over kennisorganisaties. Een meting voor deze sector bestaat niet, en dat is op zichzelf een bevinding: niemand heeft geteld hoeveel Nederlandse kennisorganisaties AI in de inhoudelijke analyse gebruiken. Het cijfer geeft een richting aan, geen sectorbeeld.

CONTROLEER WAT EEN PERCENTAGE MEET

Rond AI-adoptie circuleren percentages die iets anders meten dan ze lijken te zeggen. Neem de 77 procent uit hetzelfde LayerX-onderzoek: dat is geen adoptiecijfer, maar het aandeel gebruikers dat bedrijfsgegevens in een AI-prompt plakt. Wie zo'n getal overneemt als het aandeel medewerkers dat AI gebruikt, citeert de bron verkeerd. De controlevraag is steeds dezelfde: wat is er gemeten, bij wie, en door wie. AANNEMELIJK

Voor een kennisorganisatie weegt daar iets bij dat in veel andere sectoren minder zwaar telt: het materiaal zelf. Interviewdata en vertrouwelijke beleidsstukken bevatten vrijwel altijd persoonsgegevens. Een transcript in een privé-account is daarmee niet alleen een kwaliteitsrisico maar ook een AVG-kwestie. De organisatie blijft verwerkingsverantwoordelijke, ook wanneer één onderzoeker op eigen initiatief handelde.

Voor wie die verantwoordelijkheid draagt, is de vraag of er AI wordt gebruikt te klein geworden. De bruikbare vraag is waar en hoe dat gebeurt, en of iemand dat achteraf kan laten zien.

HOOFDSTUK 2

Welke regels gelden en wie ziet daarop toe?

De AI-wetgeving wordt intussen strenger. De invoering verloopt stapsgewijs, zonder een enkele harde deadline. En welke regels van toepassing zijn, hangt sterk af van het soort gebruik en de risicoklasse. Omdat er veel onjuiste informatie rondgaat, loont een helder overzicht juist nu.

Twee verplichtingen uit de Europese AI-verordening (Verordening (EU) 2024/1689) raken elke kennisorganisatie nu al. De geletterdheidsplicht loopt sinds februari 2025: elke organisatie moet kunnen aantonen dat medewerkers die met AI werken, weten wat ze doen. De transparantieplicht geldt sinds 2 augustus 2026, samen met de handhavingsbevoegdheid van de nationale toezichthouders. Het kader hieronder geeft de volledige kalender. BEWEZEN

De Digital Omnibus

De Digital Omnibus, het pakket dat de Raad van de Europese Unie en het Europees Parlement op 29 juni 2026 formeel aannamen, stelt de invoering van de zwaarste verplichtingen voor hoog-risico AI uit: van augustus 2026 naar eind 2027 en 2028.

Maar Artikel 4 en Artikel 50 blijven staan. Wie hieruit afleidt "het is uitgesteld, ik heb tijd", leest het verkeerd: precies de plichten die de meeste organisaties raken, geletterdheid en transparantie, blijven op schema. BEWEZEN

AI-VERORDENING · WAT GELDT WANNEER

2 FEB 2025 · VAN KRACHT

Artikel 4 (AI-geletterdheid) & Artikel 5 (verboden praktijken)

Loopt nu al. Een organisatie moet aantoonbaar AI-geletterd zijn.

2 AUG 2026 · VAN KRACHT

Artikel 50 (transparantie) + handhavingsbevoegdheid

Nationale toezichthouders kunnen vanaf dat moment handhaven, ook op Artikel 4.

2 DEC 2026

Einde overgangstermijn watermerking bestaande systemen

2 DEC 2027 · UITGESTELD

Hoog-risico AI (stand-alone, Annex III)

2 AUG 2028 · UITGESTELD

Hoog-risico AI (ingebed, Annex I)

Bron: Verordening (EU) 2024/1689 en de Digital Omnibus (Raad van de EU, 29 juni 2026).

Het eerste handhavingsmoment kwam uit een beroepsgroep

Het eerste echte Nederlandse AI-incident kwam niet van een toezichthouder, maar uit de advocatuur. In februari 2026 legden Nederlandse dekens drie advocaten een dekenale waarschuwing op wegens ondeugdelijk gebruik van generatieve AI in processtukken; twee van hen moesten verplicht een AI-cursus volgen. De aanleiding: rechters signaleerden verwijzingen naar niet-bestaande jurisprudentie. AANNEMELIJK

Let op de precieze aard. Het bleef bij deze waarschuwingen, de mildste vorm van toezicht; tot formele tuchtrechtelijke uitspraken van een Raad van Discipline kwam het niet. AANNEMELIJK

Ook in die milde vorm blijft de les staan: de eigen beroepsgroep en de opdrachtgever toetsen eerder dan de wetgever.

HOOFDSTUK 3

Geen enkel kader voor kennisorganisaties eist traceerbaarheid

Kennisorganisaties in beleidsonderzoek en evaluatie werken binnen een compact net van branchespecifieke kwaliteitsafspraken. Toch eist geen van die kaders dat een organisatie per conclusie of aanbeveling vastlegt waar die vandaan komt.

Neem de vier kaders die overal gelden.

- **ISO 20252** is een kwaliteitsnorm voor onderzoeksorganisaties, ontworpen rond de vastlegging van methode, uitvoering en analyse, zodat het onderzoek herhaalbaar blijft.
- De **Gedragscode voor Onderzoek en Statistiek**, de branchecode van MOA, VBO en VSO, gaat over integer omgaan met data, privacy en gegevensbescherming, en noemt AI überhaupt niet.
- De **Regeling periodiek evaluatieonderzoek** bepaalt wanneer beleid moet worden geëvalueerd en eist methodologische kwaliteit en onafhankelijke toetsing.
- De **Toolbox Beleidsevaluaties** van het ministerie van Financiën eist zelf niets: het is een hulpmiddel dat opdrachtgevers en evaluatoren door vijf stappen loodst, van context en scope tot rapporteren en communiceren, en zo helpt te voldoen aan de kwaliteitseisen van de RPE.

Al die kaders draaien om het proces: is het onderzoek reproduceerbaar, met dezelfde methode en dezelfde data? De opdrachtgever stelt een andere vraag: die verwijst naar een aanbeveling en wil weten waar die op gebaseerd is.

> Een onderzoek reproduceren en een bevinding herleiden zijn twee verschillende dingen.

Zo ontstaat er een gat tussen wat in de kwaliteitsvisie staat en wat de organisatie kan aantonen. Traceerbaarheid is daar een uitgangspunt, geen mechanisme. Er is niets ingericht dat het uitgangspunt afdwingt, en achteraf kan niemand de herkomst nog reconstrueren.

De academische wereld kent wel één kader met een echte herleidbaarheidseis, de gedragscode wetenschappelijke integriteit, met een herziening rond AI die in de maak is en in 2026 wordt verwacht. Die code geldt voor wetenschappelijk onderzoek bij kennisinstellingen, niet voor een onafhankelijke kennisorganisatie. De branche zal dus zichzelf moeten reguleren.

HOOFDSTUK 4

Vrijwel iedereen gebruikt AI, bijna niemand heeft het geregeld

De regels en de druk lopen vooruit op waar de organisaties in de meeste gevallen achterlopen. Die kloof is het echte probleem, niet het gebruik zelf.

Een enquête van Berenschot en Waag Futurelab onder ongeveer 500 respondenten laat zien dat ruim 90 procent in de nabije toekomst meer AI-gebruik verwacht binnen de eigen organisatie, terwijl minder dan 30 procent afspraken of beleid heeft. Laat staan een systeem.

BEWEZEN

Zet die cijfers eens naast elkaar: bijna iedereen verwacht meer AI-gebruik, minder dan een derde gaat er bewust en gericht mee om. Dat is geen teken van onzorgvuldigheid, maar van organisaties die nog geen werkbaar startpunt kregen aangereikt.

En waar wel beleid ligt, heeft dat een duidelijke vorm. Binnen die groep gaan de richtlijnen vooral over:

- privacy, gegevensbescherming en het voldoen aan wetgeving (53 procent);
- algemene gebruiksregels (51 procent).

Afspraken over inkoop, samenwerking met AI-leveranciers en medezeggenschap zijn schaars. En werkende controlemechanismen ontbreken volgens de onderzoekers bij de meeste organisaties. AANNEMELIJK

Het beleid dat er is, regelt dus vooral wat er in een AI-systeem mag. Wie controleert wat eruit komt voordat het in een rapport belandt, is ook bij de voorlopers zelden vastgelegd. Over die tweede helft gaat de rest van dit document.

Er speelt een tweede kloof die dieper zit en die minder vaak wordt benoemd. In kwalitatief evaluatieonderzoek staat de onderzoeker vaak al onder druk van de opdrachtgever.

Socioloog Vasco Lub noemde het in 2011 op Sociale Vraagstukken al een publiek geheim dat onderzoekers geregeld onder druk worden gezet om kritische passages minder negatief te formuleren, vaak doordat een opdrachtgever verlangt dat men constructief blijft. AANNEMELIJK

Die druk is niet nieuw. Nieuw is hoe geruisloos eraan kan worden toegegeven: een model zwakt een scherpe formulering in seconden af tot een acceptabele, waar dat vroeger een gesprek en een geweten kostte. Van die afzwakking blijft niets zichtbaar.

II

DEEL II

Het echte risico

HOOFDSTUK 5

Een verzonnen bron ziet eruit als een echte

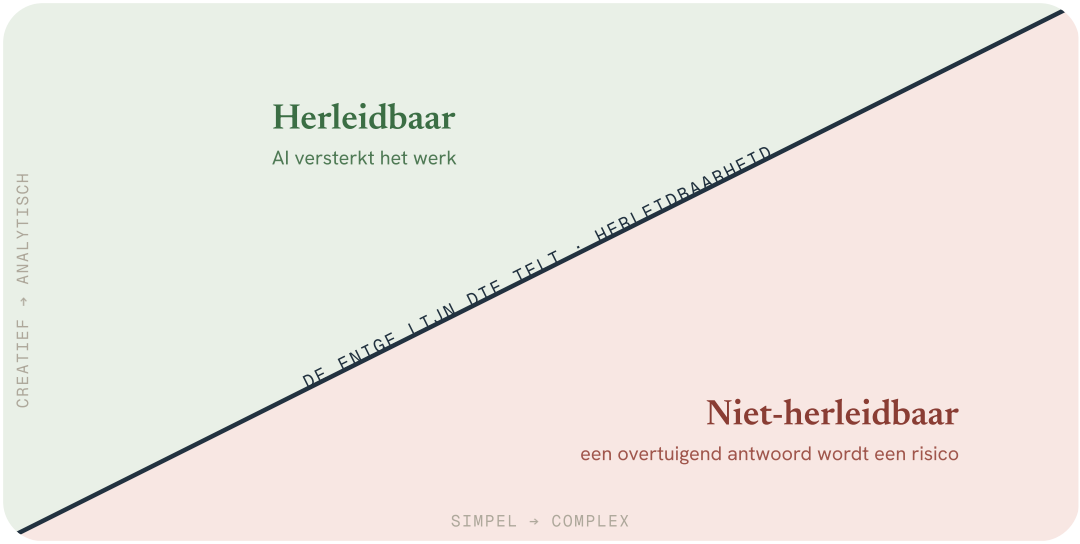
De gevaarlijkste eigenschap van AI in onderzoek is niet dat het fouten maakt. Alle gereedschappen maken fouten. Het gevaarlijke is dat deze fouten er uitstekend uitzien.

Een onderzoek onder 758 consultants (Dell'Acqua e.a., *Organization Science*, 2026) maakte beide kanten zichtbaar. BEWEZEN

- **Waar AI hielp:** creatief en schrijvend werk voor een fictief schoenenmerk, van productideeën en marktsegmentatie tot wervende teksten en memo's. De kwaliteit ging er ongeveer 40 procent op vooruit en het werk ging een kwart sneller.
- **Waar AI schaadde:** een taak die op evaluatiewerk lijkt, een advies aan de directie op basis van spreadsheetcijfers en interviews die elkaar tegenspraken. AI-gebruikers scoorden 19 procentpunt lager dan collega's zonder AI, en de kans dat een fout antwoord werd overgenomen nam juist toe.

De scheidslijn tussen AI die waarde toevoegt en AI die schaadt loopt daarom niet tussen creatief en analytisch werk, en ook niet tussen simpel en complex. De scheidslijn loopt tussen herleidbaar en niet-herleidbaar.

DE SCHEIDSLIJN DIE ERTOE DOET



De gangbare assen doen er niet toe. De enige lijn die telt, snijdt er dwars doorheen.

De vraag is of iemand die het rapport controleert, van elke bevinding, conclusie en aanbeveling kan nagaan waar die vandaan komt. Waar dat kan, is AI een krachtig hulpmiddel. Waar dat niet kan, is een overtuigend antwoord juist gevaarlijker dan een zwak antwoord, want het glipt daardoor makkelijk langs de controles.

HOOFDSTUK 6

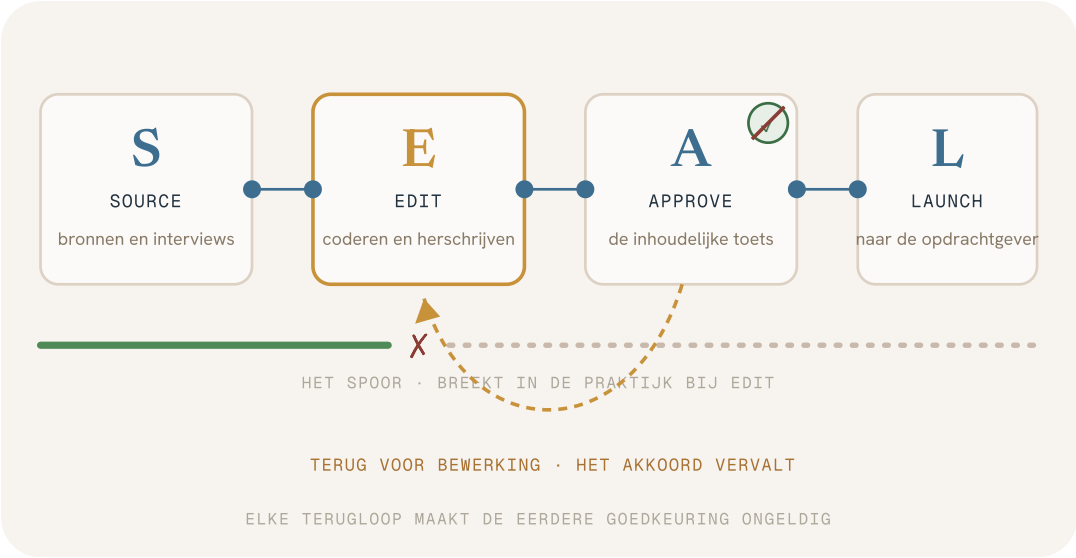
SEAL: de vier fasen waarin herleidbaarheid verloren gaat

Om aan te wijzen waar in het proces een aanbeveling losraakt van de bron, gebruiken wij een vast model, SEAL.

> SEAL beschrijft de vier fasen die kennisproductie doorloopt: **Source, Edit, Approve, Launch**. In elke fase kan de herleidbaarheid van een bron verloren gaan, en de zwakste fase bepaalt wat het werk waard is.

Bij een evaluatierapport is Source het verzamelen van bronnen en interviews, Edit alles wat daarna met het materiaal gebeurt tot aan de toets, dus ook het coderen, analyseren en samenvatten, Approve de inhoudelijke toets, en Launch het moment waarop het rapport naar de opdrachtgever gaat. Het werk loopt daarbij niet één keer door de vier fasen: een concept gaat na een toets terug voor bewerking, en elke keer dat dit gebeurt, maakt dit de eerdere goedkeuring ongeldig.

SEAL · VIER FASEN, ÉÉN TERUGLOOP



Het werk loopt niet één keer van links naar rechts: na elke toets gaat een concept terug voor bewerking, en daarmee vervalt het akkoord dat er al lag.

In elke fase kan het misgaan, maar niet elke fase is even risicovol. Waar het spoor in de praktijk breekt, staat in het volgende hoofdstuk.

HOOFDSTUK 7

De onbewaakte stap zit tussen bevinding en aanbeveling

Het risico verdeelt zich niet gelijkmatig over de vier fasen.

- **Het transcript** is te controleren: leg het naast het origineel en loop de afwijkingen na.
- **De bevinding** ook: die verwijst naar passages en documenten die in het archief liggen.
- **De aanbeveling** niet. Die ontstaat uit interpretatie, en er ligt geen brontekst naast ter controle.

Precies daar strijkt een taalmodel onder tijdsdruk een zwak onderbouwde conclusie glad tot vloeiend advies dat nergens naar terugleidt. En het is de aanbeveling waarop de opdrachtgever, de begeleidingscommissie of de toezichthouder doorvraagt.

Daar loopt de vraag stuk. De bevinding is niet verzonnen; de verbinding tussen bevinding en advies is alleen nooit vastgelegd. In SEAL-termen: **de onbewaakte stap zit bij Edit.** Bij het redigeren wordt een zin samengevat, afgezwakt of samengevoegd met een andere, en de verwijzing naar de passage waar hij op rust gaat daarbij verloren. In het eindrapport is niet meer te zien dat dit is gebeurd.

De volgende fase zou dat moeten opvangen, en juist daar zit een blinde vlek. In de praktijk is Approve een collega of projectleider die het concept doorleest: klopt de opbouw, staat er niets geks, is dit te verdedigen richting de opdrachtgever. Dat is nuttig en essentieel, maar **het is lezen, niet nalopen.**

Of een uitspraak werkelijk op de interviews en dossiers rust, is niet uit de tekst zelf te halen. Daarvoor moet bij elke conclusie het materiaal erbij worden gehaald, en daar is in een leesronde geen tijd voor. Het akkoord geldt bovendien meestal voor het rapport als geheel: de leesronde toetst het geheel, terwijl de vraag van de opdrachtgever straks over een specifieke conclusie of aanbeveling gaat.

Zo overleeft de breuk uit Edit de toets. Een akkoord dat die vraag wel aankan, ziet er anders uit: de meelezer krijgt per conclusie het materiaal ernaast, en ergens staat vast wie welke versie heeft goedgekeurd.

HOOFDSTUK 8

Ervaring beschermt niet tegen een vloeiende fout

In het voorjaar van 2026 werd een oud-hoofdredacteur van NRC en De Standaard geschorst als fellow bij Mediahuis. In 15 van 53 blogposts stonden citaten die de betrokkenen nooit hadden uitgesproken; zeven van de geciteerden ontkenden de uitspraken. BEWEZEN

Hij bevestigde zelf hoe het was gegaan: de citaten ontstonden bij het samenvatten van rapporten met AI. Het overkwam hem niet uit onwetendheid, maar door routine: samenvattingen die net niet goed genoeg werden nagelopen.

Dit voorval is geen uitzondering. In januari 2026 liep GPTZero, een bedrijf dat AI-tekst en verzonnen bronnen opspoort, 4.841 goedgekeurde papers van NeurIPS na, de grootste AI-conferentie ter wereld, en vond 100 verzonnen verwijzingen in ruim 50 papers. BEWEZEN

- Een deel was **volledig verzonnen**: niet-bestaande auteurs, een verzonnen titel, een link die nergens heen leidt.
- Een deel was **geraffineerder**: een kwalitatieve paper met verzonnen auteurs, of een verzonnen positie binnen een soms verzonnen organisatie.

Elk van die papers was beoordeeld door drie tot vijf vakgenoten, en de conferentie had een beleid dat verzonnen citaties verbiedt. Het glipte er toch doorheen, bij mensen die de technologie zelf bouwen en al maximaal waren opgeleid en geselecteerd. Beter opleiden of strenger selecteren is dus niet het antwoord.

Door beide gevallen loopt dezelfde rode draad: **werk dat af lijkt, wordt niet meer gecontroleerd.** De controle gaat na of een verwijzing er goed uitziet, niet of die bestaat. Of een bron echt is, blijkt pas wanneer iemand hem opent, en juist bij overtuigend werk gebeurt dat niet vanzelf. Dat moet het proces afdwingen.

HOOFDSTUK 9

Waakzaamheid slijt, procesinrichting niet

De voor de hand liggende maatregelen zijn beter opletten en mensen trainen. Dat helpt, maar het lost het probleem niet structureel op. En daar is bewijs voor.

In een gerandomiseerde, gecontroleerde trial (NEJM AI, 2025) kregen 44 artsen eerst 20 uur AI-geletterdheidstraining. Toch haalden de artsen die een fout AI-advies te zien kregen maar 73,3 procent diagnostische nauwkeurigheid, tegen 84,9 procent in de groep met correct advies: een verschil van 11,6 procentpunt, oplopend tot 14,0 procentpunt na correctie voor covariaten. BEWEZEN

Getraind, alert, en toch volgden ze het foute advies.

MEETCONTEXT

De fouten werden in dit onderzoek bewust in het AI-advies geïnjecteerd. Het meet dus automation bias, de neiging om een machineadvies te volgen, niet de dagelijkse foutmarge van het model.

Het mechanisme is menselijk. Bij een deskundig beredeneerd stuk lezen we vloeiendheid en logica als competentie, en zakt twijfel weg. Training dekt de wettelijke plicht tot geletterdheid, en dat is nuttig, maar het neutraliseert deze reflex niet.

De afstand van die artsen tot de onderzoekspraktijk is klein. Een ervaren onderzoeker die een conceptbevinding uit een taalmodel voorgelegd krijgt, staat er net zo voor: de AI-geletterdheidscursus is afgerond, de wettelijke verplichting afgedekt, en toch verzwakt het eigen oordeel zodra het voorstel fout maar vloeiend is.

Hoofdstuk 5 liet hetzelfde patroon zien: bij een taak waarin de bronnen elkaar tegenspraken, presteerden consultants met AI onder het niveau van collega's zonder. In evaluatieonderzoek is dat geen uitzondering maar de kern van het vak: interviews spreken elkaar tegen, dossiers spreken de interviews tegen, en juist daar maakt expertise het verschil.

Organisaties die verantwoord en optimaal gebruik willen maken van AI, onderschatten dit in veel gevallen. De training zelf is goed en nuttig: ze dekt de wettelijke plicht en leert mensen wat de systemen wel en niet kunnen. Maar wie daarna stopt, neemt een risico.

Zonder doordacht AI-gebruik, systemisch of beleidsmatig, dat bevindingen en bronnen aan elkaar gekoppeld houdt, trekt AI het niveau omlaag. Ook bij diegenen die de cursus serieus hebben gevolgd. De tijdwinst is reëel. De prijs ook: conclusies die slechter zijn dan wat hetzelfde team zonder AI leverde.

> Oplettendheid slijt. Een vaste regel niet.

Het verschil tussen oplettendheid en een vastgelegde regel gaat tellen naarmate het volume groeit. Wie honderden rapporten per jaar nakijkt, mist er vroeg of laat een. Niet uit onkunde, maar omdat niemand duizend keer achter elkaar scherp is. Bij die aantallen is een misser onvermijdelijk.

HOOFDSTUK 10

Het eerste incident in beleidsonderzoek is een kwestie van tijd

Het KPMG-geval uit de inleiding is de exacte spiegel van dit werk: een advieskantoor dat de eigen naam onder verzonnen bronnen zette. Een vergelijkbaar voorval bij een Nederlandse kennisorganisatie in beleidsonderzoek is nog niet bekend. Dat verzwakt het betoog niet; het versterkt het.

Kwaliteitsnormen komen in Nederland vrijwel altijd na een incident. De affaire-Stapel leidde jaren later tot de gedragscode wetenschappelijke integriteit. Open data werd norm toen onderzoeksfinancier NWO een datamanagementplan als subsidievoorwaarde stelde. Het patroon herhaalt zich, en het eerste publieke geval in deze hoek is een kwestie van tijd.

Twee adviseurs van Necker, Lucyenne van de Coevering en Marieke Oprel, legden in mei 2026 in het vakblad Binnenlands Bestuur uit wat er op het spel staat. Democratie draait op verantwoording: van elk besluit moet je kunnen nagaan waarop het rust. Beleidsonderzoek levert juist de rapporten en analyses waarop bestuurders en politici hun besluiten baseren. Het zit dus niet aan de rand van dat proces, maar in de kern ervan. BEWEZEN

Zodra AI die stukken meeschrijft zonder dat herleidbaar is waar een conclusie op rust, breekt de keten van verantwoording precies op dat punt. En een taalmodel kan die verantwoording niet overnemen: het is niet gekozen, het kan niet ter verantwoording worden geroepen en het draagt geen politieke verantwoordelijkheid. Die blijft bij de mens.

De verantwoordelijkheid blijft dus waar die altijd lag. Alleen het bewijs van die verantwoordelijkheid is moeilijker geworden.

BINNEN 1 MINUUT HET ANTWOORD

De opdrachtgever belt en vraagt waar een bepaalde aanbeveling op rust.

De medewerker opent niet het archief maar het rapport zelf, in de digitale omgeving waarin het is gemaakt. Bij de aanbeveling staan precies de bevindingen waarop deze is gebaseerd. Bij elke bevinding staan de passages waarop ze rust: fragmenten uit interviews en documenten, elk met een verwijzing naar de volledige bron. Waar AI is gebruikt, is terug te halen welke tekst erin ging, wat eruit kwam en wat de onderzoeker daarvan heeft overgenomen of aangepast. En bij het geheel staat wie welke versie heeft getoetst en goedgekeurd, met datum.

Het antwoord aan de opdrachtgever is dan geen reconstructie maar een uitdraai: deze aanbeveling rust op deze drie bevindingen, en die rusten op dit materiaal. Tijd tussen vraag en antwoord: minder dan een minuut. Zonder archiefonderzoek, zonder de vertrokken collega te bellen, en zonder één zin uit het hoofd te hoeven weten.

III DEEL III Aanpak

HOOFDSTUK 11

Eerst sorteren, dan pas kiezen

De meeste organisaties beginnen met de verkeerde vraag. Ze willen weten welk AI-hulpmiddel veilig is. De bruikbare vraag is: **welke documenten mogen überhaupt met AI verwerkt worden?**

Want zodra AI in beeld komt, valt informatie in twee soorten die om een tegengestelde aanpak vragen: documenten met persoonsgegevens, en kennis zonder namen. Bijna alle zorgen horen bij het eerste deel. Bijna alle veilige winst zit in het tweede.

DE SORTERING IN TWEE GROEPEN

Per document: staat er een naam in, of iets waarmee iemand te identificeren is?

1 · PERSOONSGEGEVENS

Dossiers, interviewdata, klantstukken, persoonsgegevens uit beleidsstukken.

Gaat voorlopig niet in AI, tot de juridische grondslag, de beoordeling of de anonimisering op orde zijn.

2 · KENNIS ZONDER NAMEN

Methodiek, normenkaders, gepubliceerd beleid, protocollen, eigen rapporten zonder persoonsgegevens.

Hier zit de veilige AI-winst. Begin hier.

Bij twijfel over een document geldt altijd groep 1. De twijfel is het antwoord.

Wie deze twee als één geheel behandelt, wordt gedwongen de strengste maatregelen toe te passen. De uitkomst is dan meestal dat er praktisch niets meer kan en mag, of dat er een te groot en complex systeem komt dat niemand onderhoudt.

Wie eerst sorteert, richt het deel zonder persoonsgegevens goed in en stelt het gevoelige deel uit tot de juridische basis op orde is. Die sortering kost een half uur en bepaalt de hele verdere aanpak.

HOOFDSTUK 12

Drie maatregelen die geen geld kosten

Voor de vraag welk systeem er gebouwd of gekocht wordt, zijn er drie dingen die vandaag al kunnen, niets kosten en geen aanschaf vereisen. Samen nemen ze meer echt risico weg dan welke toolkeuze dan ook.

01

Organisatie-accounts

In plaats van privé-accounts. Op een privé-account gelden consumentenvoorwaarden: de ingevoerde tekst mag vaak voor training worden gebruikt, en niemand ziet wat er wordt verwerkt. Een organisatie-account draait dat om: zakelijke voorwaarden, een verwerkersovereenkomst waar nodig, en zicht op het gebruik.

02

Tweestapsverificatie

Op elk AI-account. Zo'n account bewaart de volledige gespreksgeschiedenis, dus elk transcript en beleidsstuk dat er ooit in ging. Eén gelekt wachtwoord geeft toegang tot dat hele archief; dit is de goedkoopste manier om die deur dicht te houden.

03

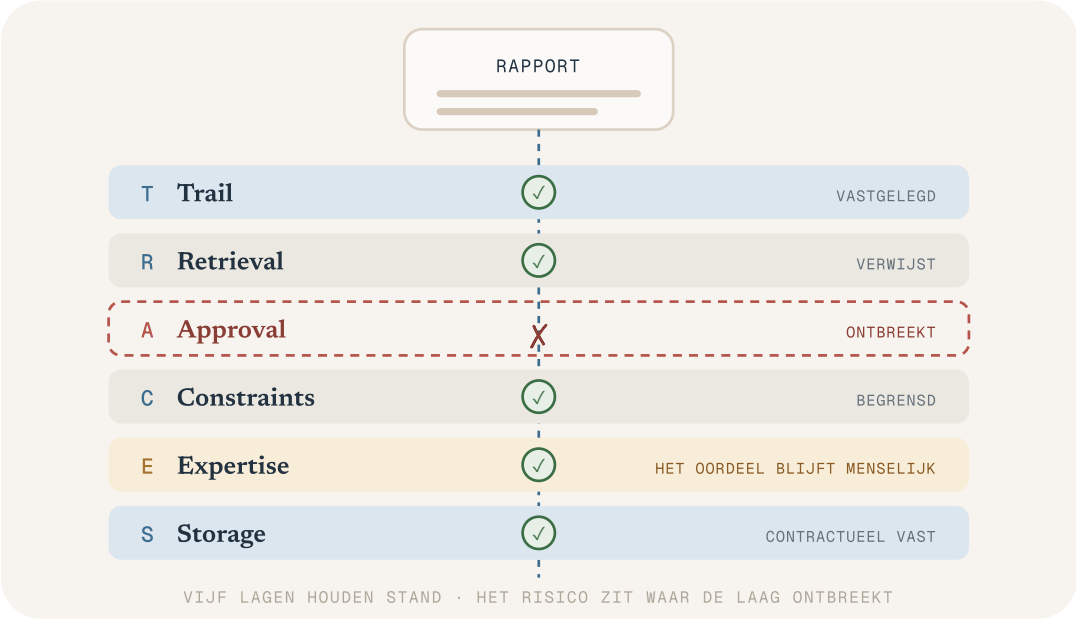
Afspraken op één A4

Geen beleidsnota, maar één A4'tje met drie antwoorden: welke gegevens gaan nooit in een AI-dienst, wie keurt goed voordat een rapport naar de opdrachtgever gaat, en bij wie meldt een medewerker een fout of twijfel.

EEN WAARSCHUWING DIE TEGEN DE INTUÏTIE INGAAT

Bij een vermoeden van verborgen AI-gebruik gaat de aandacht al snel naar de junioren. Maar het zit net zo goed bij de senior onderzoekers en de directie, en vaak in hogere mate, want daar bestaat de vrijheid om afspraken te negeren. Kijk dus naar alle lagen.

TRACES · DE DOORSNEDE



Toets een systeem laag voor laag. Vrijwel elk veelbesproken AI-incident is terug te voeren op de laag die er niet was.

HOOFDSTUK 13

TRACES: de zes lagen waarop je een systeem toetst

SEAL laat zien welke fase lekt. TRACES is waarmee je toetst of een systeem dat lek kan dichten.

Stel dus dat een organisatie iets laat bouwen of aanschaft. Hoe stel je vast of het deugt?

Een systeem dat de herkomst vasthoudt, is meer dan een model met een tekstvak eromheen. Het bestaat uit zes lagen, en een organisatie toetst bij de leverancier of elke laag er is. Ontbreekt een laag, dan zit daar een risico, en vrijwel elk veelbesproken AI-incident is terug te voeren op één of meer ontbrekende lagen.

> **TRACES** zijn de zes lagen waarop je een AI-systeem toetst: Trail, Retrieval, Approval, Constraints, Expertise, Storage. Ontbreekt een laag, dan zit daar een risico.

DE LEVERANCIERSTOETS · ZES LAGEN OP ÉÉN KAART

DE LAAG	DE VRAAG DIE JE STELT	DISKWALIFICEREND ANTWOORD
T **Trail**	Overleg het logboek van een willekeurige bevinding van vorige maand.	Iemand moest er iets voor invullen. Dan gebeurt het over drie maanden niet meer.
R **Retrieval**	Toon een bewering met bronverwijzing, én wat er gebeurt bij een vraag zonder legitieme bron.	Er komt toch iets vloeiends uit.
A **Approval**	Hoe zien de goedkeuringsstappen eruit, en bij welk deel past de beoordelaar nog iets aan?	Dat getal staat op nul: er wordt afgevinkt, niet gecontroleerd.
C **Constraints**	Welke handelingen voert het systeem zelfstandig uit, en welke liggen technisch buiten het model?	Het model schat dat zelf in.
E **Expertise**	Welke taken blijven nadrukkelijk bij mensen, en hoe voorkomt het ontwerp dat die rol uitholt?	De leverancier begrijpt de vraag niet.
S **Storage**	De verwerkersovereenkomst, de serverlocatie, de bewaartermijnen, en zwart-op-wit: geen training op jouw data.	Een verwijzing naar de algemene voorwaarden.

Zes lagen, zes vragen, en per vraag het antwoord dat einde oefening betekent. De uitwerking per laag volgt hieronder.

T Trail

Als er iets misgaat, en ergens gaat altijd iets mis, is de eerste vraag altijd dezelfde: wat is er precies gebeurd?

De Trail-laag legt bij elke bevinding vast welke bron is gebruikt, wat de conclusie was, wie heeft goedgekeurd en wanneer. Dat is niet om medewerkers te controleren, maar om het werk te verantwoorden.

De valkuil is een logboek dat iemand met de hand moet bijhouden. Onder tijdsdruk sneuvelt elke extra handeling als eerste, dus zo'n logboek wordt na twee weken niet meer ingevuld. Een goede Trail-laag vraagt daarom niets extra's: het schrijft mee en legt vast als bijproduct van het werk, en elke handeling in het systeem laat vanzelf een spoor achter.

Ontbreekt deze laag, dan verandert elk serieus incident van een herstelbare fout in verlies van vertrouwen. Een opdrachtgever of toezichthouder krijgt dan niet te zien of het om één geval gaat of om een patroon, en of de rest van het werk wel deugt.

DE VRAAG AAN DE LEVERANCIER

Overleg het logboek van een willekeurige bevinding van vorige maand. En: moest iemand daarvoor iets invullen? Zo ja, dan gebeurt het over drie maanden niet meer.

R Retrieval

Een taalmodel geeft altijd een antwoord, ook als het de nodige kennis mist. Het verzint dan iets dat klopt qua vorm, logica en toon, en juist daardoor moeilijk op fouten af te vangen is. De Retrieval-laag keert dat om:

- het systeem mag alleen antwoorden uit een afgebakende, door de organisatie beheerde bronverzameling;
- elk antwoord verwijst naar de plek waar het vandaan komt; en als de bron er niet is, meldt het systeem dat in plaats van iets te verzinnen.

Er is een tweede eigenschap die minder bekend is en die er bij onderzoek toe doet. Zoeken levert altijd iets op: het geeft de best passende treffers terug, ook als er niets werkelijk relevants is. Wat het niet kan, is aantonen dat iets ontbreekt. Welke uitspraak geen bron heeft, of welke aanbeveling nergens naar terugleidt: dat zijn negatieve vragen, en een systeem dat alleen zoekt beantwoordt die nooit. Daarvoor moet vastliggen welke bevinding aan welke bron hangt, als een vastgelegd verband en niet als een zoekopdracht.

DE VRAAG AAN DE LEVERANCIER

Laat een bewering zien met bronverwijzing, én wat het laat zien bij een vraag zonder legitieme bron. Zegt het dan eerlijk dat het antwoord er niet is, of komt er toch iets vloeiends uit? Dat tweede is diskwalificerend, hoe goed de rest ook is. En daarna: toont het systeem welke uitspraken in dit rapport geen bron hebben?

A Approval

Alles wat naar de opdrachtgever gaat, hoort eerst door een persoon te worden bevestigd. Dat klinkt vanzelfsprekend, maar de Approval-laag zit subtieler in elkaar dan een klik op een bevestigingsknop.

Een goedkeuring die uit één klik bestaat, verwatert binnen weken tot afvinken. De uitweg zit niet in meer ogen maar in minder controlemomenten. Een bron toevoegen is simpel en hoeft nergens op te wachten. Wie een uitspraak aanpast, krijgt een automatische controle: spreekt

de nieuwe formulering iets tegen dat elders in het rapport staat? Wie een conclusie of aanbeveling wijzigt, kan niet verder zonder het akkoord van de beoordelaar. De hoofdredacteur met de verzonnen citaten had geen kwade wil. De controle was een routine geworden.

DE VRAAG AAN DE LEVERANCIER

Hoe zien de goedkeuringsstappen eruit, wat krijgt de gebruiker op dat moment te zien, en welke wijzigingen worden automatisch goedgekeurd? En één getal uit het systeem: bij welk deel van de goedkeuringen past de beoordelaar nog iets aan? Wie echt controleert, vindt af en toe iets. Staat dat getal op nul, dan wordt er niet gecontroleerd maar afgevinkt.

C Constraints

Een taalmodel kan één ding: tekst maken. Het kan een zin schrijven die klinkt alsof er gerekend is, zonder dat er gerekend is, of melden dat alle namen uit een stuk zijn verwijderd, zonder dat elke naam weg is. Aan de zin zelf is dat niet te zien. Daarom horen deze handelingen, rekenen, anonimiseren, publiceren, niet af te hangen van wat het model zelf inschat.

De Constraints-laag legt gevoelige handelingen buiten het model, in vaste regels. Een model dat zelf besluit een citaat te anonimiseren, of een steekproefpercentage narekent zonder rekenmodule, of een conceptversie naar een gedeelde map schrijft, doet iets waarvan achteraf niemand kan vaststellen of het klopte. Het verschil met controle achteraf is essentieel. Een controlemoment kan worden genegeerd. Een grens die technisch vastligt niet. Dat het systeem iets niet kan, geeft meer zekerheid dan welke controle achteraf dan ook.

DE VRAAG AAN DE LEVERANCIER

Welke handelingen voert het systeem zelfstandig uit, en welke liggen technisch buiten het model?

E Expertise

Deze laag is niet technisch, maar een ontwerpkeuze, en misschien wel de belangrijkste van de zes. Een systeem kan zo worden ingericht dat het routinewerk overneemt en het oordeel bij de gebruiker laat, en dat oordeel wel of niet aanscherpt. Het kan ook zo worden ingericht dat het ongemerkt het kritische denken overneemt.

Wat dat betekent, is per rol verschillend.

- Een **junior** die nooit meer zelf een eerste analyse maakt, leert het vak anders, en niet per se beter.

- Een **projectleider** die de synthese aan een model overlaat, bouwt het eigen oordeel niet verder uit vanaf dat punt.
- Een **onderzoeker** die niet meer zelf kritisch leest, verliest voeling met de eigen expertise.

Deze laag laat zich bovendien nooit volledig automatiseren. Een systeem stelt vast dat een bron een uitspraak dekt. Een systeem stelt niet vast dat een bron de uitspraak weliswaar dekt, maar de lezer op het verkeerde been zet, doordat er selectief is geciteerd of de dekking is opgerekt. Dat vraagt vakkennis. Wel wordt het werk aanzienlijk overzichtelijker: het systeem toont bij elke uitspraak de passage waarop die steunt, zodat de beoordelaar uitspraak en passage naast elkaar legt in plaats van de hele bron door te nemen.

DE VRAAG AAN DE LEVERANCIER

Welke taken neemt het systeem over, en welke taken blijven nadrukkelijk bij mensen? En hoe voorkomt de configuratie of architectuur dat de menselijke rol langzaam uitholt? Een leverancier die deze vraag niet begrijpt, heeft er niet over nagedacht, en dat is een antwoord op zich.

S Storage

Deze laag gaat over de vraag waar informatie heen gaat en wat ermee gebeurt. Waar staat de data, in welk land, bij welke partij? Wordt er door de leverancier, of diens leverancier, getraind op wat gebruikers invoeren? Wat gebeurt er met de gegevens als het contract afloopt? En zijn de antwoorden op die vragen contractueel vastgelegd, of berusten ze op een toezegging in een verkoopgesprek?

Voor een organisatie die bijvoorbeeld met interviewdata werkt is dit geen kwestie van voorkeur: zonder verwerkersovereenkomst is de verwerking niet rechtsgeldig, wat er verder ook goed geregeld is.

DE VRAAG AAN DE LEVERANCIER

De verwerkersovereenkomst zelf, de locatie van de fysieke servers, de bewaartermijnen, en zwart-op-wit dat er niet op de gegevens van de organisatie wordt getraind. Wie hier vaag blijft of naar algemene voorwaarden verwijst die dit niet goed afdekken, valt af.

HOOFDSTUK 14

Consistentie is meetbaar

Voor een branche waar reproduceerbaarheid een kwaliteitseis is, ligt hier het sterkste stuk gereedschap. Een systeem is niet stabiel omdat de leverancier dat zegt, maar omdat een meting het uitwijst. Stel tien keer dezelfde vraag en tel hoe vaak hetzelfde antwoord met

dezelfde bronverwijzing verschijnt. Negen of tien van de tien is stabiel; zes van de tien betekent dat het antwoord mede van het toeval afhangt. Dat getal hoort thuis in de onderzoeksverantwoording, naast het responspercentage: een kwaliteitsgetal dat de lezer vertelt hoe betrouwbaar het gereedschap was waarmee het onderzoek is gemaakt.

Een systeem dat op dezelfde vraag drie verschillende aanbevelingen geeft, dient niet als onderbouwing, hoe goed elk los antwoord ook oogt. Een systeem dat tien van de tien keer naar dezelfde passage verwijst, wel.

DE PROEF, STAP VOOR STAP

1. Kies één vraag die het systeem in het echte werk moet aankunnen, bijvoorbeeld: waar rust deze aanbeveling op? Bewaar de exacte formulering.

2. Stel precies die vraag tien keer, elk in een nieuw gesprek, zodat het systeem niet voortbouwt op het vorige antwoord.

3. Noteer per keer twee dingen: het antwoord in één zin, en de bron waarnaar het systeem verwijst.

4. Tel. Hetzelfde antwoord met dezelfde bronverwijzing telt als treffer. Een ander antwoord, of hetzelfde antwoord met een andere bron, telt niet.

5. Schrijf het resultaat als breuk op: acht van de tien. Dat getal gaat de onderzoeksverantwoording in. Een half uur werk.

Tellen is geen oordeel: wie de proef herhaalt, komt op hetzelfde getal uit. De leverancier is er niet bij nodig: iedereen kan de proef vandaag zelf uitvoeren. En de proef meet precies wat de opdrachtgever stilzwijgend aanneemt: dat het onderzoek morgen hetzelfde antwoord geeft als vandaag.

DE MEETMETHODE

Vijf echte onderzoeksvragen. Elke vraag tien keer, in een schoon gesprek. Noteer hoe vaak de bevinding én de bron gelijk zijn.

Onder de acht op tien rapporteren wij een waarschuwing. Dat is geen bestaande norm maar onze keuze, gebaseerd op de gedachte dat één op de vijf afwijkende antwoorden in een rapport niet uit te leggen valt. Een andere grens mag, mits die vooraf vastligt en niet achteraf.

HOOFDSTUK 15

Eerst meten, dan kiezen

Voordat de vraag gesteld kan worden of er iets wordt gekocht of gebouwd, moet er eerst een ongemakkelijk feit onder ogen worden gezien. En het beïnvloedt de manier waarop de meeste organisaties beslissen of AI werkt of niet: op gevoel.

In een gecontroleerd onderzoek onder ervaren ontwikkelaars (METR, 2025) verwachtten de deelnemers dat de inzet van AI ze ongeveer een kwart sneller zou maken. Achteraf rapporteerden ze dat ze inderdaad het idee hadden dat het sneller ging. De meting, met de klok ernaast, liet het omgekeerde zien: ze waren bijna 20 procent langzamer met AI. En ook na afloop bleven ze geloven dat het sneller ging. De kloof tussen veronderstelde en gemeten productiviteit bedroeg zo'n 40 procentpunt. BEWEZEN

MEETCONTEXT

Het ging om 16 ervaren ontwikkelaars in door henzelf geschreven codebases, met de AI-tools van begin 2025. De onderzoekers waarschuwen zelf tegen te brede generalisatie, maar het mechanisme is niet aan het vak gebonden.

Bij onderzoek is de kans op dit effect eerder groter dan kleiner. Met AI staat er meteen iets: een concepttekst, een samenvatting, een analyse. Wat niemand registreert, is de tijd die daarna weglekt in controleren, herschrijven en bronnen natrekken, en juist dat nawerk is het grootste deel.

Een onderzoeker die een AI-concept drie keer herschrijft, een projectleider die elke bevinding naloopt, een eindredacteur die de bronnen controleert: ze ervaren allemaal de snelheid van de eerste versie en vergeten de uren erna.

Dat maakt zelfrapportage een onbetrouwbaar meetinstrument. "Ons team ervaart AI als een grote tijdwinst" is geen bewijs. Wie werkelijk wil weten of AI helpt, meet voor de invoering en meet daarna hetzelfde nog eens. Een week turven is beter dan welke enquête dan ook.

HOOFDSTUK 16

Kopen, maken of geen van beide

Koopt een kennisorganisatie een kant-en-klaar hulpmiddel, of laat ze iets maken?

Het eerlijke antwoord: kopen wint meestal. Voor routinewerk dat overal hetzelfde is, transcriptie en vertaling, is een goed bestaand hulpmiddel bijna altijd goedkoper en sneller dan zelf laten bouwen.

Maar zodra bronplicht, vertrouwelijk materiaal en de eis van herleidbaar werk samenkomen, schieten de standaard hulpmiddelen tekort. En in beleidsonderzoek is dat geen uitzondering: elk rapport dat naar de opdrachtgever gaat, moet kloppen én te verdedigen zijn. Dat is het werk.

Daar wint een andere vorm: de organisatie koopt de AI-capaciteit in en laat de laag eromheen op maat maken, het zoeken in de eigen bronnen, de bronvermelding, de goedkeuring. Dat is waar het op aankomt, en precies wat een generiek hulpmiddel niet levert.

En dan de derde uitkomst, die net zo serieus is als de eerste twee: soms is het antwoord geen van beide. Een simpel script zonder AI, een beter proces, of helemaal niets. Wijst een grondige analyse uit dat AI niets toevoegt, dan is dat een even waardevol antwoord als bouwen.

Wat alle drie de routes gemeen hebben, is de volgorde: **eerst de analyse, dan pas de keuze.** Niet andersom. Iets kopen of bouwen en dan zoeken waar het voor dient, is precies waar de meeste mislukte AI-trajecten beginnen.

HOOFDSTUK 17

De vraag die de opdrachtgever gaat stellen

Alles in dit document loopt uit op één moment: de opdrachtgever, de commissie of de toezichthouder vraagt waar een aanbeveling op rust. Wie dat moment doorstaat, verkoopt geen nauwkeurigheid maar iets waardevollers: de onderzoeksleider staat ergens voor in een zaal waar niemand het rapport zelf heeft nagerekend.

Wat een organisatie daarvoor moet tonen, leveren de zes lagen precies op: de bron naast de bevinding, de brug naar de aanbeveling, en het logboek van wie wat wanneer goedkeurde. Die inrichting komt voor de vraag, want achteraf reconstrueert niemand de herkomst nog.

> De opdrachtgever betaalt niet voor nauwkeurigheid, maar voor de zekerheid dat elke conclusie navolgbaar is zonder dat zelf na te hoeven gaan.

Dat is wat herleidbaarheid een organisatie oplevert: niet een keurmerk erbij, maar het verschil tussen een rapport dat een zaal kritische professionals overleeft en een rapport dat daar niet tegen bestand is.

En er is een tweede kant, die in dit soort discussies zelden wordt genoemd. Dezelfde inrichting die de vraag van de opdrachtgever aankan, maakt ruimte in het werk zelf: wat aantoonbaar op materiaal rust, hoeft niet uit voorzorg drie keer te worden nagelopen en mag sneller door. Het risico en de winst komen uit dezelfde afspraak.

SLOT

Afronding

SLOT

Wat deze whitepaper niet doet

Dit document geeft geen juridisch advies. Of een bepaalde datastroom is toegestaan, en of er een gegevensbeschermingseffectbeoordeling (DPIA) nodig is, is een vraag voor de functionaris gegevensbescherming of een jurist. Wat dit document doet, is de vragen aanreiken waarmee dat gesprek zinvol wordt.

Het belooft ook niet dat het risico verdwijnt. Het wordt beheersbaar, niet nul. Ook met alles op orde blijven er risico's bestaan, en daar hoort een afspraak bij, in plaats van de illusie dat het is dichtgetimmerd.

En het verkoopt niets: wie er zelf mee verder wil, heeft alles wat nodig is. De grens die wij hardop trekken: systemen die mensen beoordelen, selecteren of toelaten vallen onder het zwaarste regime en horen bij gespecialiseerde partijen, niet bij ons.

VERGELABS

Wat wij doen

Dit document verkoopt niets: wie er zelf mee verder wil, heeft alles wat nodig is. En voor wie het liever samen oppakt, is VergeLabs er.

Wij werken in één volgorde: eerst kijken waar in een organisatie echt tijd of geld weglekt, en pas daarna bouwen of adviseren. Soms is de uitkomst dat een kleinere ingreep al genoeg is, of dat AI niets toevoegt. Ook dat vertellen wij, want een goed antwoord is meer waard dan een verkocht project.

Dezelfde afspraak die het risico wegneemt, maakt de winst vrij. Na zo'n traject kun je laten zien waarop elk oordeel rust, en wat er daardoor sneller mag.

01 · Advies en analyse

Wij brengen in kaart wat met AI mag en loont, en welke documenten beter buiten een AI-dienst blijven.

02 · Training en AI-geletterdheid

Wij leren medewerkers hoe de systemen werken en wat veilig gebruik betekent. Dat dekt meteen de wettelijke geletterdheidsplicht af.

03 · Bouw van systemen die de bron vasthouden

Wij maken systemen die tonen waar een bevinding op rust en de beslissing bij de onderzoeker laten. De zes TRACES-lagen zijn daarbij de harde voorwaarde.

04 · Inrichting op de werkvloer

Wij regelen de afspraken, de zakelijke accounts en de controle, zodat AI-gebruik geregeld en zichtbaar wordt in plaats van een schaduwpraktijk.

De eerste stap is meestal gewoon een goed gesprek.

Eén zin terug is genoeg: hallo@vergelabs.nl.
Of plan een kennismaking op vergelabs.nl.

SCAN · VERGELABS.NL

Verantwoording per bewering

Elke feitelijke bewering draagt een label: BEWEZEN AANNEMELIJK OMSTREDEN. Waar een cijfer staat, staat de meetcontext erbij. De labels zijn conservatief toegekend: bij twijfel het lagere label. Waar dit document een keuze maakt die geen norm is, staat dat erbij, onder meer bij de consistentiegrens in hoofdstuk 14. Vijf

beweringen zijn op 17 augustus 2026 als steekproef opnieuw tegen de primaire bron gelegd; die dragen hieronder een verificatiedatum.

AI-gebruik onder werknemers, ~45 procent · LayerX, Enterprise AI & SaaS Data Security Report 2025 · telemetrie in browsers, geen enquête; de 77 procent is het aandeel gebruikers dat bedrijfsgegevens in een prompt plakt, geen adoptiecijfer · leverancier van browserbeveiliging AANNEMELIJK

Zelfde orde van grootte, onafhankelijke meting · Verizon, Data Breach Investigations Report 2026 · ruim 850.000 gelogde gebeurtenissen, drie keer zoveel als een jaar eerder AANNEMELIJK

>90 procent verwacht meer AI-gebruik, <30 procent heeft afspraken; beleid vooral privacy (53 procent) en gebruiksregels (51 procent) · Berenschot en Waag Futurelab, AI-Trendonderzoek 2025 · n≈500, Nederlandse organisaties breed, niet sectorspecifiek · nagelopen 17-8-2026; editie 2026 verschijnt medio september BEWEZEN

Druk van opdrachtgevers op formuleringen · Vasco Lub, Sociale Vraagstukken, 2011 · essay van een vakgenoot, geen meting AANNEMELIJK

758 consultants: ±40 procent hogere kwaliteit en een kwart sneller binnen de grens, 19 procentpunt lager erbuiten · Dell'Acqua e.a., *Organization Science*, 2026 (jrg. 37, nr. 2; eerder Harvard Business School working paper, 2023) · gerandomiseerd veldexperiment · nagelopen 17-8-2026 BEWEZEN

44 artsen, 20 uur training: 84,9 naar 73,3 procent bij fout advies, verschil 14,0 procentpunt na correctie · NEJM AI, 2025 · gerandomiseerde trial; de fouten waren bewust geïnjecteerd, het onderzoek meet automation bias BEWEZEN

Ontwikkelaars ~19 procent langzamer met AI, terwijl ze dachten sneller te zijn · METR, 2025 · gerandomiseerd, 16 ervaren ontwikkelaars in eigen codebases; de onderzoekers waarschuwen zelf tegen brede generalisatie BEWEZEN

NeurIPS: 100 bevestigde verzonnen verwijzingen in 51 tot 53 papers · GPTZero, rapport 21 januari 2026 · 4.841 van 5.290 goedgekeurde papers nagelopen; elk paper had drie tot vijf reviewers · primaire bron staat open, nagelopen 17-8-2026 BEWEZEN

Deloitte Australië betaalde deel van overheidsopdracht terug · najaar 2025, rapport voor het ministerie van werkgelegenheid; verzonnen academische bronnen en een verzonnen citaat uit een rechterlijke uitspraak · internationale persberichtgeving, o.a. Financial Times en AFR BEWEZEN

KPMG trok AI-adviesrapport terug; 5 van 45 bronverwijzingen klopten · GPTZero-analyse, juni 2026, van het rapport "Redefining excellence in the age of agentic AI" (oktober 2025) · Financial Times; ook The Register, TechCrunch en accountant.nl · nagelopen 17-8-2026 BEWEZEN

EY Canada (mei 2026) en PwC Middle East (vier rapporten, eind juli 2026) · persberichtgeving, PwC via GPTZero in de Financial Times · nog niet tegen de primaire artikelen gelegd AANNEMELIJK

Drie advocaten kregen een dekenale waarschuwing; twee moesten verplicht een AI-cursus volgen · NOS, 22 februari 2026 · het bleef bij de mildste vorm van toezicht, geen tuchtuitspraken AANNEMELIJK

Oud-hoofdredacteur van NRC en De Standaard geschorst: verzonnen citaten in 15 van 53 blogposts, zeven geciteerden ontkennen · NRC, 19 maart 2026; NOS, 20 maart 2026 · betrokkene bevestigde zelf dat de citaten ontstonden bij het samenvatten van rapporten met AI · nagelopen 17-8-2026 BEWEZEN

Verantwoording als kern van beleidsonderzoek · Van de Coevering en Oprel (Necker), Binnenlands Bestuur, mei 2026 · duiding door vakgenoten, geen meting BEWEZEN

De wettelijke kalender · Verordening (EU) 2024/1689 en de Digital Omnibus (aangenomen 29 juni 2026) · geletterdheid sinds 2 februari 2025; transparantie en handhaving sinds 2 augustus 2026; overgangstermijn watermerking tot 2 december 2026; hoog-risico uitgesteld naar 2 december 2027 en 2 augustus 2028 BEWEZEN

Geraadpleegde kaders zonder cijferclaims · ISO 20252 · Gedragscode voor Onderzoek en Statistiek (MOA, VBO en VSO) · Regeling periodiek evaluatieonderzoek · Toolbox Beleidsevaluaties · gedragscode wetenschappelijke integriteit, herziening rond AI verwacht in 2026.

Bijlage · wat vandaag al kan

Zonder ons, zonder budget en zonder tool. Wie alleen deze bijlage uitvoert, staat verder dan de meeste organisaties.

1 · Sorteer in twee groepen: een half uur

Loop je materiaal langs: staat er een naam of herleidbaar persoon in? Zo ja of bij twijfel: groep 1, voorlopig niet in AI. Zo nee: groep 2, hier zit je veilige winst.

2 · Drie gratis maatregelen: een middag

Organisatie-accounts in plaats van privé, tweestapsverificatie op elk AI-account, en één A4 met afspraken: wat nooit in AI, wie tekent, waar meld je een fout.

3 · Stel de zes TRACES-vragen

Neem ze mee naar het eerstvolgende leveranciersgesprek: Trail, Retrieval, Approval, Constraints, Expertise, Storage. Werkt de leverancier met een demo, stel ze daar dan ook.

4 · Meet een week: voor je iets vindt

Turf een week lang de tijd van een taak mét en zónder AI, vóórdat iemand een mening vormt. Gevoel is geen bewijs; een gemeten week wel.