

HANDLEIDING

AI-MEETPLAN

VOOR HET SUCCESVOL TOEPASSEN VAN AI

VOOR

Elke organisatie die met AI werkt
of AI wil gaan gebruiken

DATUM

15 september 2026

OPGESTELD DOOR

Scavian B.V.

VERSIE

1.0

VRIJ TE GEBRUIKEN EN TE VERSPREIDEN

Colofon

TITEL	AI-meetplan
ONDERTITEL	Voor het succesvol toepassen van AI
AUTEUR	Scavian B.V.
VERSIE	1.0
DATUM	16 september 2026
STATUS	Definitief

Gebruik

Dit document mag je vrij gebruiken en verspreiden. Je hoeft ons (Scavian B.V.) niet om toestemming te vragen als je een deel of de hele inhoud wilt overnemen. Je hoeft zelfs de bron niet te vermelden.

Feedback

Er is altijd ruimte voor verbetering. Ontbreekt er iets, klopt er iets niet, of werkt het in jouw situatie heel anders? Laat het weten. We stellen je feedback op prijs en verwerken aanvullingen in een volgende versie. Met naamsvermelding, als je wilt.

Disclaimer

We moeten wel één voorbehoud maken. Scavian B.V. is niet verantwoordelijk voor enige schade als gevolg van het gebruik van dit document.

Contact

ORGANISATIE	Scavian B.V.
TELEFOON	+31 33 200 3939
E-MAIL	hallo@scavian.ai
WEBSITE	scavian.nl
KVK	98031287
BTW	NL868331880B01

Versiebeheer

VERSIE	DATUM	AUTEUR	WIJZIGING
0.1	10 september 2026	Scavian B.V.	Eerste concept
0.2	12 september 2026	Scavian B.V.	Herziene opzet in zeven stappen; meetplan, enquête en toets als losse documenten; concept voor praktijktest
0.3	14 september 2026	Scavian B.V.	Redactionele wijzigingen
0.4	15 september 2026	Scavian B.V.	Redactionele wijzigingen
1.0	16 september 2026	Scavian B.V.	Finale wijzigingen voor publicatie

Inhoudsopgave

Colofon	2
Inhoudsopgave	4
Inleiding: waar het AI-meetplan voor dient	5
Over het AI-meetplan	6
Stap 1: de AI-toepassing	7
Stap 2: doelen en eigenaars	8
Stap 3: opdrachten en risico's	10
Stap 4: nulmeting	12
Stap 5: meetpunten en streefwaarden	13
Stap 6: wat mag niet slechter worden	15
Stap 7: evaluatiemomenten	16
Tot slot	17
BIJLAGE A CIJFERS VERZAMELEN	18
BIJLAGE B MEETPUNTEN EN WAT NIET SLECHTER MAG WORDEN	21
BIJLAGE C ENQUÊTE ONDER MEDEWERKERS	25
BIJLAGE D DE TOETS	26
Contact	27

Inleiding: waar het AI-meetplan voor dient

Donderdagmiddag, vier uur. De evaluatie van de AI-pilot staat op de agenda. Er is taart. De projectleider vertelt dat de medewerkers enthousiast zijn en dat er zeker een vervolg komt op deze pilot. Iedereen knikt, niemand vraagt door en om kwart over vier wordt verdergegaan met het volgende agendapunt.

De meeste AI-projecten in organisaties beginnen zonder vooraf vastgelegde doelstelling. Er is niets afgesproken waaraan je het resultaat kunt toetsen, dus blijft alleen het enthousiasme van de voorvechters over. Dat enthousiasme is meestal groot, want de meest betrokken mensen laten zich het duidelijkst horen. Het is een wankel basis voor een soms grote verandering in de manier waarop een organisatie werkt.

Het enthousiasme van de koplopers weerhoudt de minder enthousiaste medewerkers ervan zich uit te spreken. Valt de AI-toepassing tegen, dan stoppen zij niet met AI. Ze kijken uit naar eigen oplossingen, zoals een privéabonnement, buiten het zicht van de organisatie.

Het AI-meetplan helpt beide problemen voorkomen. Het helpt je vast te leggen wat je met AI wilt bereiken en te meten of dat lukt: beoordeling op cijfers, niet op indrukken. Het meet ook wat een systeem niet over zichzelf kan meten. Of medewerkers er echt beter werk door leveren, of ze de antwoorden controleren en of ze toch andere AI gebruiken. Daarvoor maakt een enquête onderdeel uit van het meetplan (bijlage C).

Het AI-meetplan is bedoeld voor iedereen die een besluit over het toepassen van AI moet nemen of verantwoorden. Het is niet gebonden aan een leverancier of een product. Het werkt als AI nog niet formeel is geïntroduceerd, als je AI breed hebt uitgerold onder alle medewerkers, als één team een AI-toepassing gebruikt (gekocht of zelfgebouwd) en als je een AI-toepassing voor klanten of bezoekers overweegt.

Het AI-meetplan maakt duidelijk wat je met AI wilt bereiken en of je dat bereikt.

De termen in dit document

Elke vorm van AI die door een organisatie wordt gebruikt noemen we een AI-toepassing: een chatbot op de website, een hulpmiddel voor medewerkers, een programma dat zelfstandig taken uitvoert, gekocht of zelfgebouwd. Hoe de leverancier het noemt, assistent, agent of chatbot, doet er voor het meten niet toe. Wie de AI-toepassing levert of beheert noemen we de leverancier, ook als dat je eigen IT-afdeling is of als je de toepassing zelf hebt gebouwd.

Over het AI-meetplan

Een meetplan stel je op voor één AI-toepassing en één set van maximaal drie doelen. Het beantwoordt zeven vragen, in zeven stappen.

1. **De AI-toepassing.** Moeten gebruikers inloggen? Dat bepaalt wat je kunt meten.
2. **Doelen en eigenaars.** Wat willen we bereiken met de AI-toepassing en wie is daarop aanspreekbaar?
3. **Opdrachten en risico's.** Wat moet de AI-toepassing doen, wat mag ze niet doen en hoe erg is een fout?
4. **Nulmeting.** Waar staan we nu?
5. **Meetpunten en streefwaarden.** Waaraan zien we straks dat we bereiken wat we willen? Wanneer grijpen we in als dat niet wordt bereikt?
6. **Wat mag niet slechter worden.** Welke bijwerkingen accepteren we niet, ook niet als de doelen worden gehaald?
7. **Evaluatiemomenten.** Wanneer beoordelen we, wie is betrokken en wie besluit?

Wie zit er aan tafel

Het invullen van het meetplan kost een dagdeel als de juiste mensen aan tafel zitten: de eigenaren van de doelen, iemand die de inhoud kent en resultaten kan beoordelen, iemand die de cijfers van de organisatie kan leveren en iemand die weet wat de AI-toepassing rapporteert, vaak de leverancier of de eigen IT-afdeling. Dus niet alleen de projectleider en de voorvechters. Duurt het langer dan een dagdeel, dan staan er waarschijnlijk te veel doelen op de lijst.

Eén meetplan per AI-toepassing

Heeft je organisatie drie AI-toepassingen, dan maak je drie meetplannen. Een nieuwe set doelen voor dezelfde AI-toepassing is een nieuw meetplan, dat naar het vorige verwijst. Elk meetplan krijgt een kenmerk dat binnen de organisatie uniek is, bijvoorbeeld de naam van de AI-toepassing, het jaar en een volgnummer: HR-assistent-2026-01. Is de AI-toepassing nog niet gekozen, gebruik dan een werknaam.

De documenten

Bij deze handleiding horen losse documenten: een leeg meetplan, een volledig ingevuld voorbeeld voor BouwKwest (een verzonden bouwbedrijf van 140 medewerkers), een enquête voor medewerkers en de toets waarmee je gaat testen of de toepassing doet wat ervan wordt verwacht, leeg en als ingevuld voorbeeld. In deze handleiding staan bij elke stap dezelfde tabellen als in het lege meetplan, met de uitleg erbij. In het meetplan vul je bij eigenaren en betrokkenen een rol in, zodat het meetplan ook zonder namen gedeeld kan worden en bruikbaar blijft als medewerkers van rol of functie veranderen. Namen voeg je toe bij de ondertekening.

Stap 1: de AI-toepassing

Niet elke AI-toepassing laat zich op dezelfde manier analyseren. Eén vraag bepaalt wat je kunt meten: moeten gebruikers inloggen?

Als gebruikers moeten inloggen, dan zijn ze binnen de toepassing identificeerbaar en dus bekend. De toepassing is dan “besloten”. Je kunt dan meten welk deel van de doelgroep de AI-toepassing gebruikt (“bereik”), of gebruikers terugkomen en of ze na drie maanden nog actief zijn (“terugkeer”).

Hoeven gebruikers niet in te loggen, dan is de AI-toepassing “publiek” en zijn gebruikers anoniem. Dan zijn bereik en terugkeer niet te meten. Je bent dan aangewezen op volume en de verdeling van het verkeer.

Bekend is niet hetzelfde als gevolgd. Cijfers over gebruikers moeten binnen de AI-toepassing worden berekend en alleen als totaal of percentage worden gerapporteerd, nooit per persoon. Meer daarover in stap 5.

WAT DE AI-
TOEPASSING DOET

GEBRUIKERS
IDENTIFICEERBAAR

Ja / Nee

Als je AI-toepassing hybride is

Een AI-toepassing is hybride als het zowel een publiek als een besloten deel kent. Neem een webshop met een AI-toepassing op de site. Een niet-ingelogde bezoeker stelt vragen over producten en levertijden. Zodra diezelfde persoon inlogt en iets vraagt over zijn bestelling, is bekend wie de bezoeker is. Stel in zo’n situatie een meetplan op voor het publieke deel en een ander meetplan voor het besloten deel. Vraag je leverancier of de rapportage onderscheid maakt tussen gesprekken met ingelogde en niet-ingelogde gebruikers. Eén meetplan met doelen voor beide delen zorgt alleen maar voor verwarring. Dat wil je voorkomen.

Stap 2: doelen en eigenaars

Nu je weet wat je kunt meten, leg je vast wat je met de AI-toepassing wilt bereiken. Dit is de stap die in de meeste AI-projecten wordt overgeslagen. Daardoor worden die projecten als succesvol bestempeld zonder dat iemand kan aantonen dat het werkelijk een succes is.

Kies de drie belangrijkste doelen die je wilt bereiken, in gewone taal, SMART geformuleerd en elk doel met een eigenaar. Kies je meer doelen, dan verlies je focus.

Controlevraag per doel: als dit doel over een jaar is bereikt, wat is er dan zichtbaar verbeterd? Komt daar geen concreet antwoord op, dan is het doel te vaag om te meten.

Doel 1

DOEL

EIGENAAR (ROL)

ZICHTBAAR BETER
OVER EEN JAAR

Doel 2

DOEL

EIGENAAR (ROL)

ZICHTBAAR BETER
OVER EEN JAAR

Doel 3

DOEL

EIGENAAR (ROL)

ZICHTBAAR ANDERS
OVER EEN JAAR

Wat telt als succes

Of gebruikers moeten terugkomen of juist in één keer geholpen moeten zijn, is onderdeel van het doel zelf. Een naslagtoepassing voor medewerkers wil dat mensen blijven terugkomen. Een klantenservicetoepassing wil dat een klant met dezelfde vraag niet terug hoeft te komen. Eén AI-toepassing kan doelen van beide soorten hebben. Het doel bepaalt daarmee welke meetpunten in stap 5 zinvol zijn en welke niet.

De eigenaar

De eigenaar is degene die over een jaar op het doel wordt aangesproken. In het meetplan vul je de rol in. De naam komt erbij in stap 7, bij de ondertekening. Zorg voor overdracht van het meetplan bij een wijziging in de relevante rollen of functies.

SMART formuleren

Het SMART-principe gebruik je voor het eenvoudig en eenduidig opstellen en controleren van doelstellingen. De letters staan voor:

- **Specifiek.** Is de doelstelling eenduidig?
- **Meetbaar.** Onder welke meetbare of observeerbare voorwaarden is het doel bereikt?
- **Acceptabel.** Is het doel acceptabel voor de doelgroep en het management?
- **Realistisch.** Is het doel haalbaar?
- **Tijdgebonden.** Wanneer moet het doel bereikt zijn?

Meer over SMART: nl.wikipedia.org/wiki/SMART-principe.

Stap 3: opdrachten en risico's

Een AI-toepassing hoeft niet overal even goed in te zijn, maar op sommige onderwerpen mogen geen fouten worden gemaakt. In deze stap benoem je wat de AI-toepassing moet doen en hoe erg het is als dat misgaat. Dat zorgt ervoor dat het besluit om live te gaan op cijfers berust en niet op gevoel of durf.

Opdrachten

Een opdracht is wat een gebruiker wil van de AI-toepassing: een vraag beantwoorden of een taak uitvoeren. Voorbeelden zijn vragen over verlof, een loonvaststelling opstellen of een mailbox opschonen. Benoem maximaal zes soorten opdrachten, met per soort het gewenste resultaat en de risicoklasse. Denk ook aan wat de AI-toepassing niet mag doen, bijvoorbeeld vragen beantwoorden over een persoonlijk dossier. Het gewenste resultaat is dan aangeven dat de vraag niet kan of mag worden beantwoord of dat doorverwezen moet worden naar een mens.

Risicoklassen

Het risico van een fout resultaat verschilt per soort opdracht. De eisen die je stelt, stem je daarop af.

KLASSE	KENMERK	VOORBEELD	WAT DAT VRAAGT
Laag	Een fout resultaat is hinderlijk en de gevolgen zijn omkeerbaar	Welke opleidingen bestaan er?	Normale zorgvuldigheid
Midden	Een fout resultaat kost tijd of leidt tot een verkeerde stap	Welke procedure geldt hier?	Bronvermelding verplicht
Hoog	Een fout resultaat kost geld, rechten of vertrouwen	Vragen over arbeidsvoorwaarden, inschaling, verlof; taken zoals gegevens verwijderen of mails versturen	Toets vóór livegang gehaald volgens de norm uit het meetplan

Tabel: risicoklassen

De toets vóór livegang

Je wilt testen of de toepassing doet wat ervan wordt verwacht. Deze toets stel je samen ná het invullen van het meetplan. Dit doe je met de mensen die de materie inhoudelijk kennen, bijvoorbeeld met HR-specialisten als de toepassing bedoeld is om dat team te ondersteunen.

De toets bestaat uit een set testopdrachten. Dit zijn voorbeelden per soort opdracht uit de eigen praktijk, met het resultaat dat de organisatie juist vindt. Een soort opdracht kan zijn: vragen beantwoorden die met het personeelshandboek of de cao te beantwoorden zijn. Een testopdracht is dan: op hoeveel vakantiedagen heeft een medewerker recht? Hoe je de toets samenstelt, staat in bijlage D. In het meetplan leg je nu per soort opdracht alleen de norm vast, bijvoorbeeld 90 procent van de testopdrachten voldoet. Je legt ook vast wie het resultaat beoordeelt.

Geef je de AI-toepassing opdrachten met risicoklasse hoog, dan gaat die toepassing pas live als de toets is gehaald. Doe geen concessies op dit punt. Bij klasse laag of midden geldt een minder strikte houding. Bij die risicoklassen is de toets aan te raden, niet noodzakelijk. De toets herhaal je na elke wijziging van de AI-toepassing of van de informatie waarmee die toepassing werkt.

NR	SOORT OPDRACHT	GEWENST RESULTAAT	KLASSE	NORM TOETS	BEOORDELAAR (ROL)
1					
2					
3					
4					
5					
6					

Stap 4: nulmeting

Een doel zegt waar je heen wilt. De nulmeting legt vast waar je vandaan komt. Zonder die beginsituatie kun je over een jaar wel zeggen dat het beter is geworden, maar niet laten zien hoeveel terwijl dat het gesprek is dat je wilt kunnen voeren.

Per doel leg je vast wat je nu al meet, de huidige waarde, de bron en de datum. Is er geen cijfer beschikbaar, noteer dat dan als zodanig en leg vast hoe en wanneer je alsnog meet. Een bekend gat is beter dan een verzonnen getal. Een cijfer dat je alleen kunt krijgen door twee weken te tellen, is die twee weken waard.

DOEL	WAT METEN WE NU	HUIDIGE WAARDE	BRON	DATUM
1				
2				
3				
Publiek	Bezoekers per maand van de pagina waar de AI-toepassing komt te staan			

Gaat het om een publieke AI-toepassing, voeg dan het huidige aantal bezoekers toe van de pagina waar de AI-toepassing komt te staan. Doe dat ook met de vindbaarheid op de belangrijkste zoektermen, als je website dat meet. Met die cijfers bepaal je of tegenvallend gebruik komt door de AI-toepassing of door te weinig bezoekers. Dat onderscheid bepaalt of je de toepassing of de vindbaarheid moet verbeteren.

De enquête

Als de gebruikers identificeerbaar zijn, dan kun je die gebruikers de AI-toepassing laten beoordelen via de enquête bij het meetplan. Gebruikers zijn identificeerbaar als ze moeten inloggen (besloten AI-toepassing). Ze zijn ook identificeerbaar als je de gebruikers kent die een publieke AI-toepassing gebruiken. Je weet dan dat het je medewerkers zijn die de toepassing gebruiken. Dat laatste geldt bijvoorbeeld als je medewerkers allemaal ChatGPT of Copilot (gaan) gebruiken.

Verstuur de nulmeting twee tot vier weken vóór livegang en de nameting twaalf weken erna. De enquête in bijlage C kun je direct gebruiken voor medewerkers. Pas de enquête aan als de gebruikers geen medewerkers zijn maar klanten of leden.

Bij een publieke AI-toepassing is een enquête niet uitvoerbaar als er geen bekende groep is om te bevragen. Daar is de beoordeling door gebruikers in de AI-toepassing zelf het enige instrument dat je kunt hanteren (bijlage A). Dit geldt alleen als de toepassing over deze mogelijkheid beschikt én erover rapporteert.

Stap 5: meetpunten en streefwaarden

De doelen staan, de opdrachten zijn benoemd en de beginsituatie is bekend. Nu kies je de cijfers waarmee je straks vaststelt of de AI-toepassing doet wat ervan wordt verwacht. Kies zorgvuldig. Een paar goed gekozen meetpunten geven een scherper beeld dan een dashboard vol getallen.

Kies per doel maximaal twee meetpunten en leg per meetpunt vast wie het cijfer levert, wat de streefwaarde is in week 12 en in week 26 en bij welke waarde de eigenaar van het doel ingrijpt of stopt. De letter geeft aan wie het cijfer levert: S het systeem zelf, B een beoordeling door een mens van je organisatie, O je eigen organisatie. De lijst met meetpunten staat in bijlage B; hoe je de cijfers verzamelt, staat in bijlage A.

Doel 1

MEETPUNT	WIE LEVERT	STREEF- WAARDE WEEK 12	STREEF- WAARDE WEEK 26	INGRIJPGRENS
----------	---------------	------------------------------	------------------------------	--------------

Doel 2

MEETPUNT	WIE LEVERT	STREEF- WAARDE WEEK 12	STREEF- WAARDE WEEK 26	INGRIJPGRENS
----------	---------------	------------------------------	------------------------------	--------------

Doel 3

MEETPUNT	WIE LEVERT	STREEF- WAARDE WEEK 12	STREEF- WAARDE WEEK 26	INGRIJPGRENS
----------	---------------	------------------------------	------------------------------	--------------

Streefwaarde en ingrijpgrens

Een streefwaarde is de waarde waarbij het doel op koers is: een getal met eenheid, bijvoorbeeld maximaal 28 vragen per week of minimaal 50 procent. Een ingrijpgrens is de waarde waarbij de doeleigenaar ingrijpt of stopt, met het moment waarop dat geldt. Die afspraak maak je nu, zolang er nog geen uitkomst ligt die iemand liever anders had gezien. De doeleigenaren tekenen ervoor in stap 7.

Kiezen uit de lijst

Kies uit de lijst in bijlage B, of neem een eigen cijfer van de organisatie als meetpunt (O), zoals het aantal vragen dat een afdeling per week ontvangt of de doorlooptijd van een proces. Kies meetpunten die passen bij wat het doel wil: terugkeer alleen als herhaald gebruik gewenst is, afhandeling in één gesprek alleen als dat het doel is. Neem de definitie uit de lijst over, zodat iedereen hetzelfde telt.

Cijfers over gebruikers, geaggregeerd

Bereik en terugkeer vereisen dat het systeem per gebruiker bijhoudt of hij actief was, maar niemand heeft daarvoor inzage per persoon nodig: de AI-toepassing moet de cijfers of percentages leveren over de hele groep.

Een beheeromgeving die laat zien wie wanneer inlogt en hoeveel vragen stelt, biedt hiermee de mogelijkheid om het gedrag van medewerkers te controleren. Dat is in Nederland niet zomaar toegestaan. Vraag je leverancier om cijfers zonder inzage per gebruiker en neem dat op in je eisen; zie bijlage A.

Stap 6: wat mag niet slechter worden

De stappen hiervoor gaan over wat beter moet worden. Deze stap gaat over de bijwerkingen die niet acceptabel zijn, ook niet als de doelen worden gehaald.

Leg minimaal twee punten vast, waarvan één over de kwaliteit van de resultaten, elk met een grenswaarde en met wat je doet als die wordt overschreden. De meest interessante punten zijn vaak ook vervelende punten: werkdruk, de vrijheid om een collega iets te vragen en of mensen dingen minder goed zelf opzoeken.

Denk bij elk doel en elk meetpunt na over wat er niet mag gebeuren. Als, bijvoorbeeld, ‘minder contacten met een medewerker’ een doel is voor een AI-klantenservicetoepassing, dan is dat alleen winst als de tevredenheid van de klant over die contacten niet daalt.

NR	WAT MAG NIET SLECHTER WORDEN	GRENSWAARDE	WAT DOEN WE BIJ OVERSCHRIJDING
1			
2			
3			
4			

Een keuzelijst met punten en waar het cijfer vandaan komt, staat in bijlage B.

Rapporteer deze punten bij elke evaluatie, ook als de rest er goed uitziet. Als je er niet over rapporteert, geeft de evaluatie geen eerlijk beeld.

Stap 7: evaluatiemomenten

Cijfers die niemand op een afgesproken moment bekijkt, veranderen niets. In deze laatste stap leg je vast wanneer je evalueert, wie betrokken is en wie besluit. Dat doe je nu, zolang er nog geen uitkomst ligt die iemand liever anders had gezien.

Er zijn vier vaste evaluatiemomenten. De weken in de tabel kunnen worden aangepast als jullie situatie daar om vraagt.

MOMENT	WAT WORDT BEOORDEELD	WIE IS BETROKKEN (ROLLEN)	WIE BESLUIT (ROL)
Vóór livegang	De toets is gehaald volgens de normen uit stap 3		
Week 4	Werkt de AI-toepassing technisch en zijn de cijfers uit stap 5 leverbaar en bruikbaar		
Week 12	Eerste inhoudelijke evaluatie tegen de streefwaarden uit stap 5 en de grenswaarden uit stap 6		
Week 26	Besluit over voortzetten, uitbreiden of stoppen		

Ondertekening

De doeleigenaren en de besluitnemers tekenen voordat de AI-toepassing live gaat voor de doelen, de streefwaarden en de ingrijpgrenzen. Een groep, zoals de directie of het managementteam, tekent via een vertegenwoordiger. Daarmee is het meetplan een afspraak en geen voornemen.

ROL OF GROEP	NAAM	DATUM	HANDTEKENING

Gebruik het volledig ingevulde meetplan voor de fictieve organisatie BouwKwest als voorbeeld.

Tot slot

Het invullen van dit meetplan kost je een dagdeel en geeft je de mogelijkheid om over een jaar te weten of het iets heeft opgeleverd. Daarom maak je deze afspraken vooraf en niet achteraf.

Ontbreekt er iets, klopt er iets niet, of werkt het in jouw situatie heel anders? Laat het weten; de contactgegevens staan in het colofon.

BIJLAGE A

CIJFERS VERZAMELEN

Deze bijlage werkt uit hoe je de cijfers uit stap 5 in de praktijk verzamelt en wat je van een leverancier mag vragen. De beoordeling door een mens levert de B-cijfers; de enquête en de beoordeling door gebruikers in de AI-toepassing leveren O-cijfers; de laatste paragraaf gaat over de S-cijfers die je van je leverancier mag verwachten.

Beoordelen zonder mee te lezen

Of een resultaat inhoudelijk klopt, kan geen enkel systeem betrouwbaar over zichzelf vaststellen. Daar heb je mensen voor nodig. De voor de hand liggende manier, gesprekken van gebruikers meelezen, is bij een AI-toepassing voor medewerkers geen optie: een gesprek met een AI-toepassing is persoonlijk en meelezen door de organisatie is een vorm van controle op medewerkers die alleen onder strikte voorwaarden mag. Een goede AI-toepassing laat het dan ook niet toe: de inhoud van een gesprek is alleen zichtbaar voor de gebruiker zelf. Je beoordeelt daarom op drie andere manieren.

- **De toets, periodiek herhaald.** Dezelfde testopdrachten als vóór livegang, bijvoorbeeld maandelijks opnieuw uitgevoerd door de beoordelaar uit stap 3. Dit is de zuiverste maat voor kwaliteit: bekende opdrachten met een bekend juist resultaat. Zie bijlage D.
- **Gesprekken die gebruikers zelf delen.** Een gebruiker die een resultaat beoordeelt, een fout meldt of een gesprek doorstuurt ter beoordeling, deelt dat gesprek bewust. Die gesprekken mag de beoordelaar lezen. Zorg bij besloten AI-toepassingen (waar gebruikers moeten inloggen) dat melden en delen slechts één klik kost.
- **Signalen zonder inhoud.** Het systeem kan tellen zonder te lezen: vragen waarvoor geen bron werd gevonden, herformuleringen, afgebroken gesprekken, lage beoordelingen. Die tellingen wijzen aan waar je de toets moet uitbreiden.

Bij een publieke AI-toepassing met anonieme gebruikers ligt het anders: daar is een steekproef van gesprekken misschien mogelijk, mits je onder het invoerveld meldt dat gesprekken beoordeeld kunnen worden en de gesprekken niet tot een persoon herleidbaar zijn. Ook dan zijn de gedeelde en gemelde gesprekken de waardevolste.

Twee kanttekeningen bij een steekproef van gesprekken:

- Een anonieme gebruiker kan in een gesprek informatie verstrekken die toch tot een persoon herleidbaar is. Wees dus terughoudend met deze steekproeven, ook als zo'n steekproef veel informatie over mogelijke verbeteringen kan opleveren.
- Belangrijker dan de uitkomst van de beoordeling is of gevonden fouten ook worden gemeld. Als mensen fouten wel zien maar niet doorgeven, is er werk te doen.

Beoordeling door gebruikers in de AI-toepassing

Laat een gebruiker elk resultaat beoordelen met één tot vijf sterren bij de vraag hoe tevreden hij is met dit resultaat, met een optioneel veld voor een toelichting dat pas verschijnt na de sterren. De sterren kosten maar één klik. De toelichting is waar de waardevolle informatie zit, want daarin staat de reden voor de beoordeling.

Verwacht niet dat veel gebruikers de toelichting invullen, want zodra iemand antwoord heeft op zijn eigen vraag, is de motivatie om nog iets te doen meestal verdwenen. Dat maakt de toelichting die je wel krijgt waardevol. Behandel de sterren als indicatie en de toelichtingen als leesmateriaal voor de beoordelaar. Vraag je leverancier van de AI-toepassing naar de mogelijkheden.

En houd voor ogen wat je meet. Een gebruiker kan niet altijd beoordelen of een resultaat klopt. Een overtuigend klinkend en toch fout antwoord krijgt dan misschien vijf sterren. Tevredenheid meten is daarom geen vervanging voor de toets, maar een aanvulling erop.

Metten met de enquête

Zijn je gebruikers identificeerbaar (bijvoorbeeld omdat ze inloggen), dan kun je ze bevragen met een korte, aantoonbaar anonieme enquête vóór de start en opnieuw na twaalf weken, met dezelfde vragen: de nulmeting vraagt naar verwachtingen, de nameting naar ervaringen. Laat de uitnodiging niet uitgaan van de opdrachtgever van het project en niet van een leidinggevende, maar van een afzender buiten het project en buiten de lijn: HR, de personeelsvertegenwoordiging of een staffunctie. Rapporteer geen subgroepen kleiner dan acht personen, zodat niemand individueel herkenbaar is. Blijft de respons onder de vijftig procent, dan is de uitkomst een indicatie en geen meting. De enquête voor medewerkers staat in bijlage C.

Uitwijkgedrag signaleren

In de praktijk betekent uitwijken een privéabonnement op een ander model, met bedrijfsdocumenten erin geplakt. Buiten het zicht, zonder belemmerende procedures of regels en met het risico dat de data buiten Europa staat en voor het trainen van modellen wordt gebruikt.

Ook dan kunnen de cijfers over het gebruik van je eigen AI-toepassing er redelijk uitzien, want de eenvoudige vragen stellen mensen er nog wel. De vragen die ertoe doen gaan naar een andere toepassing, waar je geen controle over hebt. De enquête bevat twee vragen die dit gedrag zichtbaar maken: gebruik je voor je werk ook AI buiten wat de organisatie aanbiedt en, zo ja, wat is daarvoor de belangrijkste reden? Die tweede vraag is de belangrijkste, want mensen wijken zelden uit vanwege beleid. Ze wijken uit vanwege kwaliteit of gedoe.

Weeg dat mee in je keuze. Een AI-toepassing die merkbaar trager of zwakker is dan wat mensen thuis gebruiken, verliest het, hoe goed het beleid ook is opgeschreven. Bezuinigen op kwaliteit is daarmee zelden zo goedkoop als het lijkt.

Wat je van je leverancier mag vragen

Welke leverancier je ook kiest, dit zijn redelijke eisen:

- Cijfers over gebruik, bronvermelding en storingsen, zichtbaar in je eigen beheeromgeving, zonder dat je erom hoeft te vragen.
- Cijfers over gebruikers als totaal of percentage, zonder inzage per gebruiker.
- Geen inzage in de inhoud van gesprekken voor beheerders of leverancier, tenzij de gebruiker een gesprek zelf deelt of meldt.
- Een voorziening waarmee gebruikers een resultaat kunnen beoordelen, een fout kunnen melden en een gesprek ter beoordeling kunnen delen, met functionaliteit waarmee de beoordelaar van de organisatie die meldingen kan lezen.
- Onderscheid in de rapportage tussen gesprekken met en zonder ingelogde gebruiker, als je AI-toepassing beide kent.
- Een duidelijk onderscheid tussen gemeten cijfers en aannames. Wordt er een tijdsbesparing gerapporteerd, vraag dan hoe die berekend is. Het aantal beantwoorde vragen vermenigvuldigd met een aangenomen aantal minuten is een aanname en geen meting.
- Inzicht in de staat van de informatie waarmee de AI-toepassing werkt: welke informatiebronnen (data, documenten) verouderd zijn en welke nooit worden geraadpleegd. Vraag ook inzicht in welke informatiebronnen wel beschikbaar zijn gesteld maar niet bruikbaar zijn binnen de AI-toepassing. Dat laatste komt vaker voor dan je zou verwachten en is van buitenaf onzichtbaar.
- Duidelijkheid over waar je gegevens staan, hoe lang ze bewaard worden en of gespreksinhoud gebruikt wordt voor het trainen van modellen.

BIJLAGE B

MEETPUNTEN EN WAT NIET SLECHTER MAG WORDEN

De lijst met meetpunten voor stap 5 en de keuzelijst met punten die niet slechter mogen worden voor stap 6. De letter geeft aan wie het cijfer levert: S het systeem, B een beoordeling door een mens van de organisatie (de toets en de door gebruikers gedeelde gesprekken), O de eigen organisatie.

Lijst met meetpunten (stap 5)

Altijd van toepassing

MEETPUNT	WIE LEVERT	DEFINITIE
Responstijd	S	Tijd tussen het versturen van een vraag en het begin van het antwoord, gemiddeld over de periode.
Aandeel antwoorden met bronvermelding	S	Antwoorden waarin de AI-toepassing een bron uit de eigen informatie van de organisatie noemt, gedeeld door alle antwoorden.
Aandeel vragen waarvoor geen bron gevonden werd	S	Vragen waarbij het systeem geen passende bron vond, gedeeld door alle vragen. Een hoog aandeel kan betekenen dat mensen vragen stellen die de organisatie nog nooit heeft beantwoord; dat is waardevolle informatie.
Aandeel gesprekken zonder antwoord door een storing	S	Gesprekken die eindigden met een technische fout, gedeeld door alle gesprekken.
Aandeel vervolgvragen dat een herformulering is	S	Vervolg vragen die dezelfde vraag anders stelden, gedeeld door alle vervolgvragen; een maat voor niet-begrepen antwoorden.
Aandeel beoordeelde antwoorden dat feitelijk klopt	B	In de beoordeling (de periodieke toets en de door gebruikers gedeelde gesprekken): antwoorden die overeenkomen met de bron, gedeeld door alle beoordeelde antwoorden.
Aandeel beoordeelde antwoorden dat de vraag volledig beantwoordt	B	In de beoordeling: antwoorden die de gestelde vraag beantwoorden en niets wezenlijks weglaten.

MEETPUNT	WIE LEVERT	DEFINITIE
Aandeel opdrachten waarin terecht naar een mens werd verwezen	B	In de beoordeling: doorverwijzingen die volgens de beoordelaar terecht waren, gedeeld door alle testopdrachten waarin doorverwezen had moeten worden.

Alleen als gebruikers geïdentificeerd kunnen worden

Alle cijfers over gebruikers worden door het systeem berekend en alleen als totaal of percentage gerapporteerd, nooit per persoon. Gebruik de terugkeercijfers alleen als herhaald gebruik bij dit doel gewenst is; bij een klantenservicedoel zijn hoge terugkeercijfers een waarschuwingssignaal.

MEETPUNT	WIE LEVERT	DEFINITIE
Aandeel van de doelgroep dat de AI-toepassing gebruikte	S	Gebruikers die in de periode minimaal één vraag stelden, gedeeld door de omvang van de doelgroep.
Aandeel gebruikers dat na de eerste week terugkwam	S	Gebruikers die in de eerste week een vraag stelden en na de eerste week nog eens, gedeeld door alle gebruikers uit de eerste week.
Aandeel gebruikers uit de startmaand dat in maand twee en drie nog actief is	S	Gebruikers die in de eerste maand een vraag stelden en in maand twee en drie nog eens, gedeeld door alle gebruikers uit de eerste maand.
Ervaren of gemeten tijdsbesparing	O	Uit de enquête (ervaren) of uit een registratie van de organisatie (gemeten).
Vertrouwen in de antwoorden	O	Uit de enquête: het aandeel gebruikers dat het eens is met de stelling dat de antwoorden kloppen.

Alleen als afhandeling in één gesprek bij dit doel gewenst is

MEETPUNT	WIE LEVERT	DEFINITIE
Aandeel vragen dat in één gesprek werd afgehandeld	S en B	Gesprekken zonder doorverwijzing en zonder vervolgesprek over hetzelfde onderwerp, gedeeld door alle gesprekken. De gebruikersbeoordeling en de toets controleren of het antwoord ook klopte.

MEETPUNT	WIE LEVERT	DEFINITIE
Aandeel gebruikers dat binnen een afgesproken termijn terugkomt met hetzelfde onderwerp	S	Alleen als gebruikers inloggen: gebruikers met een tweede gesprek over hetzelfde onderwerp binnen de termijn, gedeeld door alle gebruikers met een gesprek over dat onderwerp.

Alleen bij een publieke AI-toepassing

Een toepassing die op eigen kracht bezoekers trekt, heeft verkeer dat gelijkmatig over de dagen verdeeld is. Een toepassing die op aandacht drijft, heeft pieken rond nieuwsbrieven en andere communicatie en daartussen vrijwel niets. Dat inzicht bepaalt of je moet investeren in de kwaliteit van de toepassing of in de vindbaarheid ervan.

MEETPUNT	WIE LEVERT	DEFINITIE
Bezoekers op de pagina	O of S	Unieke bezoekers per maand van de pagina waar de AI-toepassing staat, uit de webstatistieken van de organisatie of van de leverancier.
Aandeel bezoekers dat een gesprek start	S en O	Gestarte gesprekken gedeeld door bezoekers van de pagina, per maand.
Deel van het verkeer op de vijf drukste dagen	S	Gesprekken op de vijf drukste dagen van de periode, gedeeld door alle gesprekken in de periode.

Alleen bij een AI-toepassing die taken uitvoert

Voor een AI-toepassing die taken uitvoert, bijvoorbeeld documenten opstelt of een mailbox opschooft, gelden dezelfde beginselen met andere meetpunten.

MEETPUNT	WIE LEVERT	DEFINITIE
Aandeel resultaten dat zonder correctie bruikbaar was	B	In de beoordeling (de periodieke toets en de door gebruikers gedeelde resultaten): resultaten die de organisatie zonder aanpassing kon gebruiken, gedeeld door alle beoordeelde resultaten.
Aandeel taken dat handmatig moest worden hersteld	O	Taken waarvan het resultaat door een medewerker is teruggedraaid of gecorrigeerd, gedeeld door alle uitgevoerde taken.

MEETPUNT	WIE LEVERT	DEFINITIE
Doorlooptijd per taak	O	Tijd van opdracht tot bruikbaar resultaat, inclusief controle door een medewerker, gemiddeld over de periode.
Aandeel taken dat de AI-toepassing terecht niet uitvoerde	B	In de beoordeling: testopdrachten waarin de AI-toepassing terecht stopte of terugverwees naar een mens, gedeeld door alle testopdrachten waarin dat had gemoeten.

Keuzelijst voor wat niet slechter mag worden (stap 6)

WAT MAG NIET SLECHTER WORDEN	WAAR HET CIJFER VANDAAN KOMT
Kwaliteit van de resultaten	aandeel antwoorden of resultaten dat voldoet in de periodieke toets en in de door gebruikers gedeelde gesprekken
Gemelde fouten	medewerkers die een fout resultaat zagen, hebben dat gemeld (enquête, vragen C4 en C5)
Tevredenheid van klanten of medewerkers	uit de bestaande klanttevredenheidsmeting of de enquête
Aantal contacten met een medewerker	bij een doel dat afhandeling in één gesprek beoogt: mag niet dalen zonder dat de tevredenheid gelijk blijft
Ervaren werkdruk	enquête, stelling D1
Vrijheid om een collega een vraag te stellen	enquête, stelling D2
Eigen vaardigheid: dingen zelf opzoeken en onthouden	enquête, stelling D4
Druk om de AI-toepassing te gebruiken	enquête, stelling D5
Uitwijkgedrag: gebruik van andere AI buiten wat de organisatie aanbiedt	enquête, vragen D6 en D7
Doorlooptijd van een proces	eigen registratie van de organisatie
Beheerlast voor de organisatie	uren per week voor het bijhouden van de informatie waarmee de AI-toepassing werkt en het beoordelen van resultaten

BIJLAGE C

ENQUÊTE ONDER MEDEWERKERS

De enquête meet wat een systeem niet kan meten: of AI je medewerkers echt helpt, of ze de resultaten vertrouwen en of ze toch andere AI gebruiken. De enquête wordt twee keer afgenomen, één keer vóór de introductie van de AI-toepassing en één keer twaalf weken erna, met dezelfde vragen.

De nulmeting vraagt naar de huidige situatie en naar verwachtingen, de nameting naar de nieuwe situatie en de ervaringen met de AI-toepassing. De volledige vragenlijsten staan in het losse document 'Enquête: nulmeting en nameting onder medewerkers', zodat je ze in je eigen enquêtetool kunt zetten.

Randvoorwaarden

- Echt en aantoonbaar anoniem: geen registratie van IP-adres of account, geen open vraag waarvan het antwoord iemand kan identificeren.
- De uitnodiging is neutraal geformuleerd en komt van een afzender buiten het project en buiten de lijn.
- Deelname is vrijwillig en niet zichtbaar voor leidinggevenden.
- Geen uitkomsten rapporteren voor subgroepen kleiner dan acht respondenten.
- Dezelfde vragen bij nulmeting en nameting, zodat verwachting en ervaring vergelijkbaar zijn.
- Vanaf vijftig procent respons is de uitkomst een meting; daaronder een indicatie.
- Een enquêtetool die geen persoonsgegevens vastlegt en binnen de EU host, gekozen door de organisatie en niet door de leverancier van de AI-toepassing.

Voordat je begint

Laat vóór de nulmeting drie tot vijf medewerkers de vragenlijst hardop invullen en vertel niet wat de bedoeling is. Een vraag die op twee manieren te lezen is, komt dan boven water voordat hij de uitkomst bederft. Gebruik in de vragen de naam van de AI-toepassing die iedereen kent en vul in de vragen over de huidige situatie de soorten opdrachten uit stap 3 in.

Voor klanten of leden

De vragenlijst is geschreven voor medewerkers. Kunnen gebruikers worden geïdentificeerd maar zijn het klanten of leden, dan kun je dezelfde opzet gebruiken met een kortere lijst; zie de laatste paragraaf van de enquête.

BIJLAGE D

DE TOETS

De toets is de lijst van testopdrachten met bekend juist resultaat. Hiermee toets je vóór livegang of de AI-toepassing doet wat er in stap 3 is vastgelegd en daarna stel je die kwaliteit periodiek opnieuw vast. De toets is een apart document, opgesteld na het meetplan door de mensen die de inhoud kennen. Een lege toets en een ingevuld voorbeeld horen bij deze handleiding.

Zo stel je de toets samen

- Per soort opdracht uit stap 3 meerdere testopdrachten, met de meeste bij klasse hoog. Neem ook testopdrachten op waarin de AI-toepassing moet doorverwijzen of weigeren.
- Elke testopdracht is een echte vraag of opdracht uit de praktijk, geformuleerd zoals een gebruiker hem stelt.
- Het gewenste resultaat is controleerbaar. Bij een vraag: het juiste antwoord, met de bron erbij. Bij een taak: de eisen waaraan het resultaat moet voldoen, bijvoorbeeld welke mails weg mogen en welke niet, of waaraan een ontwerp moet voldoen.
- Het gewenste resultaat wordt vastgesteld door iemand die de inhoud kent. Niet door de leverancier en niet door de AI-toepassing zelf.
- Maak de testopdrachten niet beschikbaar in de AI-toepassing. Anders toets je of het antwoord kan worden opgezocht uit de lijst met testopdrachten in plaats van of het antwoord uit de echte bronnen kan worden afgeleid.

Zo gebruik je de toets

Leg elke testopdracht voor aan de AI-toepassing en beoordeel het resultaat als 'voldoet' of 'voldoet niet' tegen het gewenste resultaat; een resultaat dat maar deels voldoet, voldoet niet. Per soort opdracht is de toets gehaald als het aandeel dat voldoet minimaal gelijk is aan de norm uit het meetplan. Herhaal de toets na elke wijziging van de AI-toepassing of van de informatie waarmee zij werkt en bewaar de uitkomsten: de toets is een blijvend bezit van de organisatie en de zuiverste maat voor kwaliteit die je hebt.

NR	TESTOPDRACHT	GEWENST RESULTAAT	BRON OF EISEN	UITKOMST
1				Voldoet wel/niet
2				Voldoet wel/niet
...				Voldoet wel/niet

Tabel: opzet van de toets

Contact

ORGANISATIE	Scavian B.V.
TELEFOON	+31 33 200 3939
E - MAIL	hallo@scavian.ai
WEBSITE	scavian.nl
KVK	98031287
BTW	NL868331880B01