

MAASSTAD ZIEKENHUIS

een santeon ziekenhuis



werkenbijmaasstadziekenhuis.nl

Boven alles
Rotterdams

MAASSTAD ZIEKENHUIS

een santeon ziekenhuis



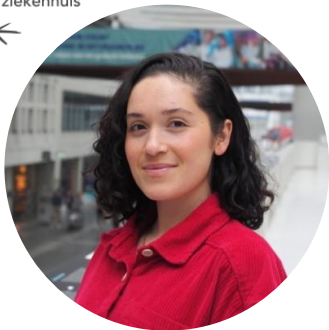
Lokale validatie van medische AI

René Verhaart en Sade Krijgsman

klinisch informatici

9 april 2026





Even voorstellen...



Sade Krijgsman

- Achtergrond in Biomedische wetenschappen (MSc)
- Klinisch informaticus sinds 2022 (EngD)
- Werkzaam in vakgroep Klinische Fysica & Informatica
- Lid medische staf
- Nevenfuncties
 - Vice-voorzitter NVKI
 - Senior AI consultant bij Romion Health

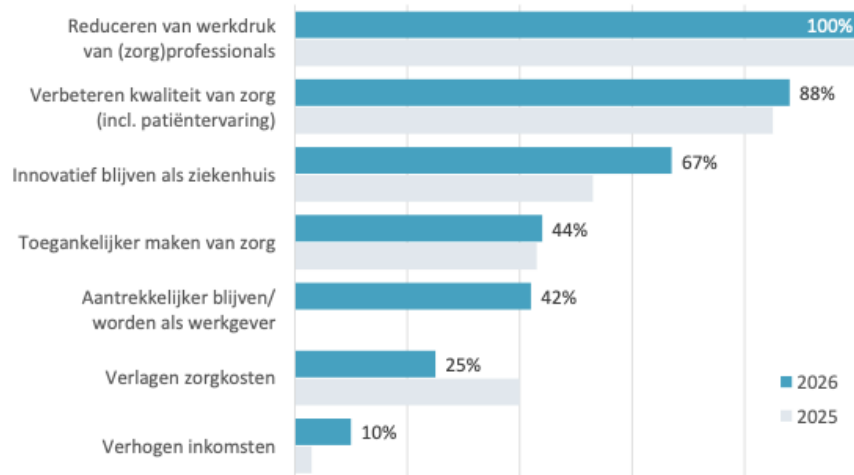
René Verhaart

- Achtergrond in Biomedische technologie (MSc en PhD)
- Klinisch informaticus sinds 2015 (EngD)
- Werkzaam in vakgroep Klinische Fysica & Informatica
- Lid medische staf
- Nevenfuncties
 - NVKI commissie 'beroepsontwikkeling & profilering'

AI in de ziekenhuizen in NL

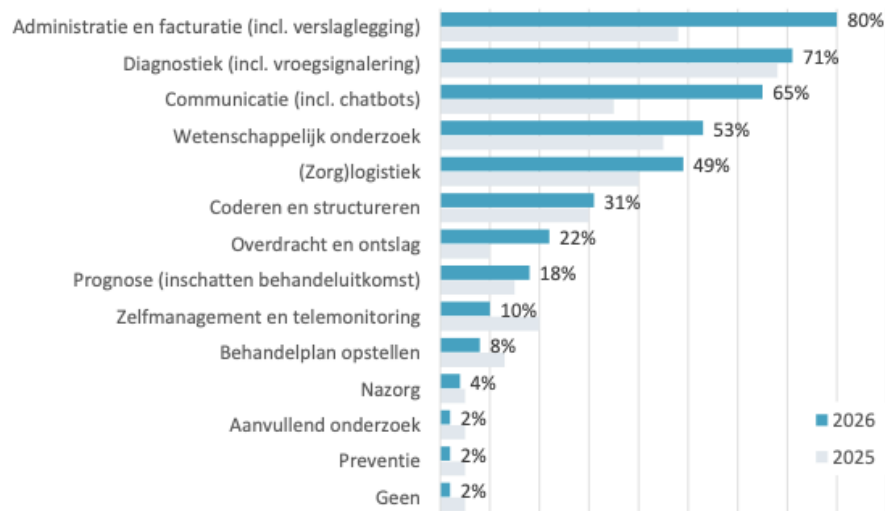
Wat zijn binnen uw ziekenhuis de voornaamste drijfveren voor het gebruik van AI?

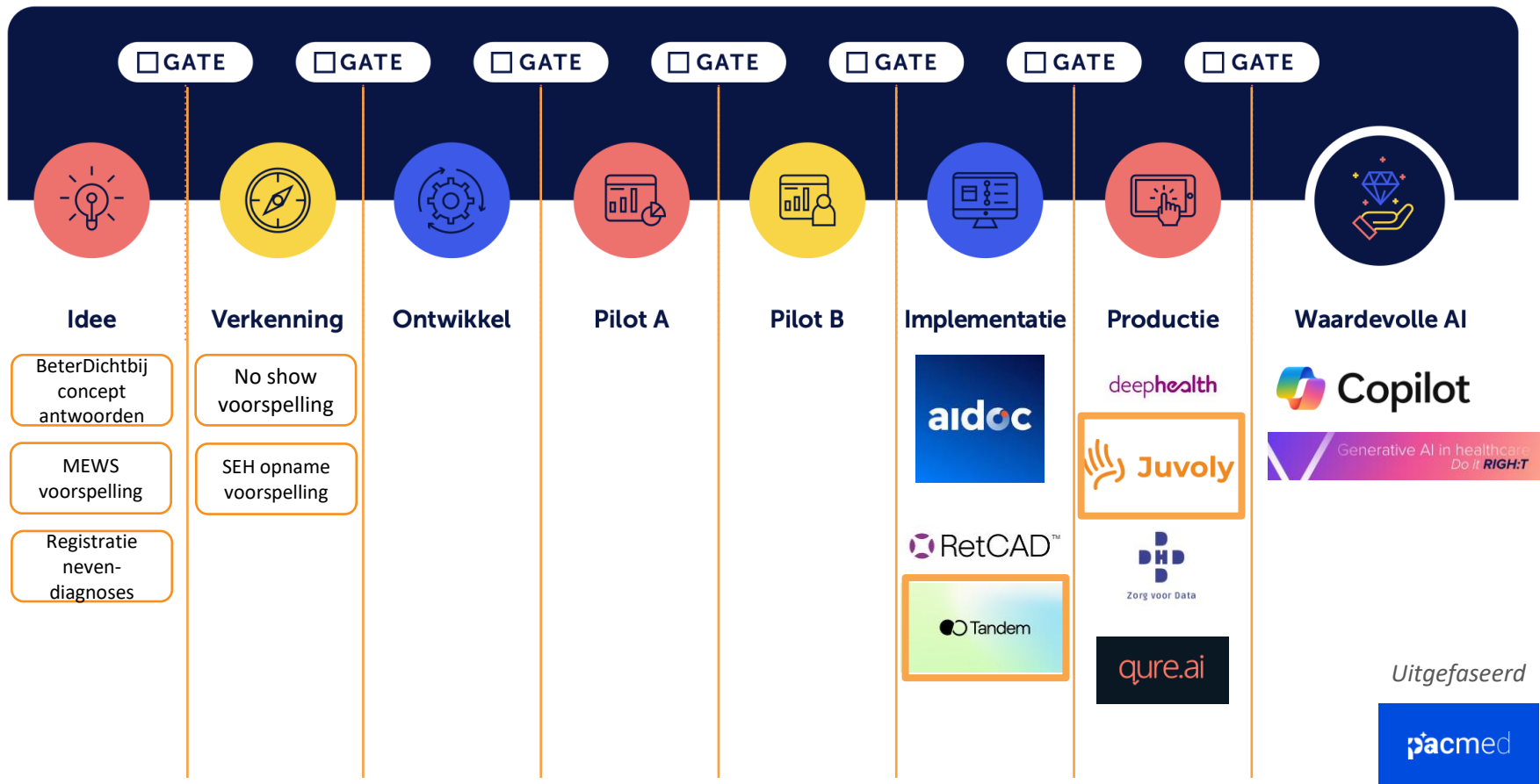
n=52 ziekenhuizen



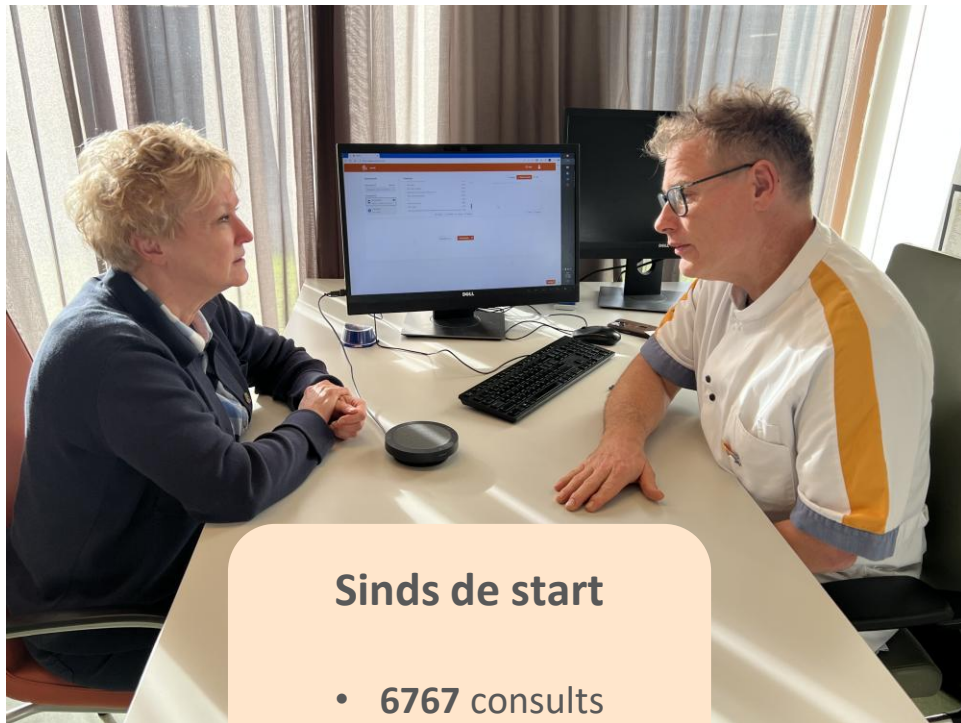
Bij welke (zorg)processen wordt AI ingezet binnen uw ziekenhuis?

n=52 ziekenhuizen





Spraakgestuurd rapporteren



Sinds de start

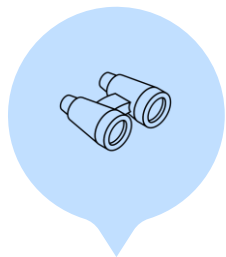
- **6767** consults
- **103** gebruikers

Uw zorgverlener luistert,
de technologie schrijft mee



Zo heeft uw zorgverlener meer
tijd en aandacht voor u

Welke stappen hebben we doorlopen?



November 2024

Markt verkenning

Welke leveranciers zijn
er op de markt?



Februari 2025

Investerings aanvraag

Projectenproces,
investeringscommissie
en RvB



April 2025

Akkoord RvB

Financiële middelen
beschikbaar



Mei 2025

Start Implementatie

Eindelijk gebruiken van
de technologie...



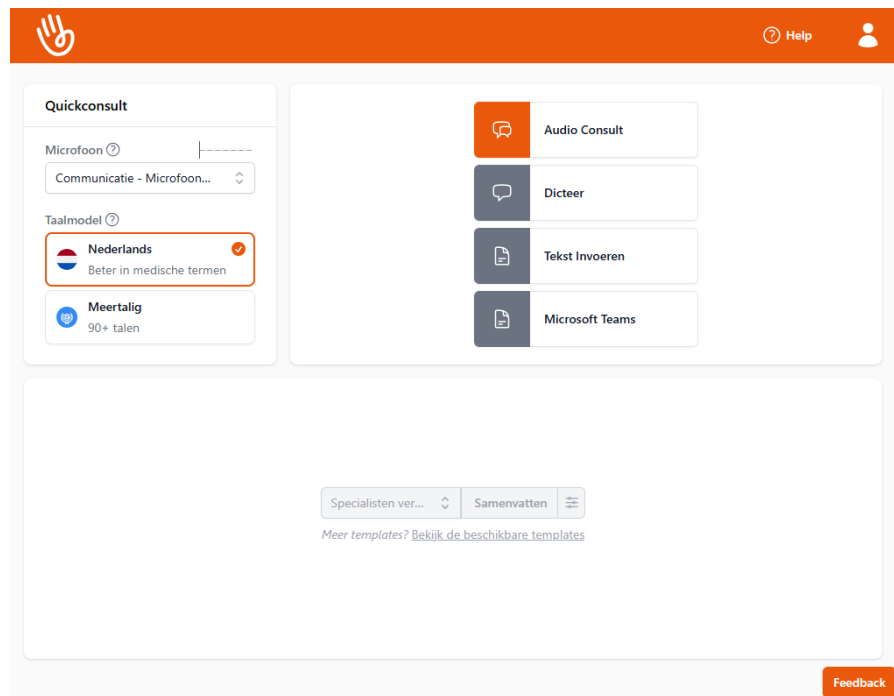
Januari 2026

In gebruik in 10 vakgroepen

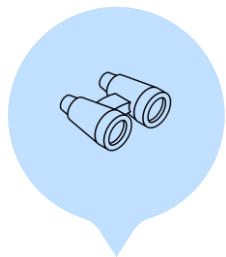
Start opschaling...

Wat is er nu in gebruik?

- PEC en PVC met voorkeuren
 - Focus op anamnese, beleid en samen beslissen
- Overige templates
 - Kindergeneeskunde
 - Neonatologie
 - Familiegesprekken (nog in testfase)
 - Ergotherapie
- Voor wie?
 - Artsen, PA en VS
 - Interne geneeskunde, reumatologie, kindergeneeskunde, Revalidatie, pijnpoli, dermatologie, chirurgie, orthopedie, mdl neurologie



Welke stappen hebben we doorlopen?



November 2024

Markt verkenning

Welke leveranciers zijn
er op de markt?



Februari 2025

Investerings aanvraag

Projectenproces,
investeringscommissie
en RvB



April 2025

Akkoord RvB

Financiële middelen
beschikbaar



Mei 2025

Start Implementatie

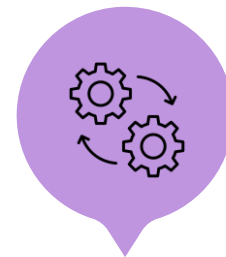
Eindelijk gebruiken van
de technologie...



Januari 2026

In gebruik in 10 vakgroepen

Start opschaling...



Komende periode

Migratie naar Tandem

(her) validatie...

January 26, 2026

Tandem Health acquires leading Dutch AI medical scribe Juvoly to scale clinician-first AI documentation across Europe

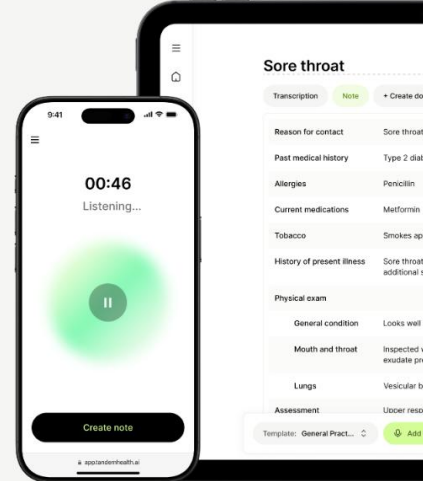


News



EU MDR
Class IIa

Tandem's Coding Assistant
is now **MDR Class IIa** certified



Intended purpose Scribe (CE klasse I)

*The Tandem AI Medical Scribe is intended for use as a **documentation and clinical consultation workflow aid** to assist healthcare professionals with dictating, transcribing and summarising consultations and meetings, including drafting downstream messaging and communications relevant to the consultation and health records. **It is not to be used to inform clinical diagnosis, treatment, or medical decision-making**. The device also offers non-medical administrative assistance to support the post-consultation actions set out by the responsible healthcare professional. All outputs generated by Tandem, including notes, summaries, and documents, must be reviewed and verified by the responsible healthcare professional prior to use.*

Interface Tandem



Vraag het Tandem

Dicteren

+ Nieuw consult

Differentiaal diagnose

Ribletsel of een ribcontusie, DD een ribfractuur

Diagnose

Bovensteluchtweginfectie

Beleid

- Paracetamol advies |
- Controle 2 weken

↶ Reset

Overwegingen

- Schone longauscultatie, pneumonie onwaarschijnlijk
- Op basis hiervan keuze voor conservatief, symptomatisch beleid

🔊

Kopiëren

Aanpassen

Disclaimer

Disclaimer: Dit verslag is mede met AI gemaakt en gecontroleerd door de zorgverlener.

Toegevoegde codes (1)

ICD-10 J06.9 Acute infectie van bovenste luchtwegen, niet gespecificeerd

✕

Voorgesteld

🔗 Vastgezette

🔍 Zoekcodes

Meest relevant

ICD-10 R05 Hoesten

?

...

Template: Specialist intake ...

+ Item toevoegen

Aanpassen

Kopiëren

MAASSTAD
ZIEKENHUIS
een tandheel ziekenhuis

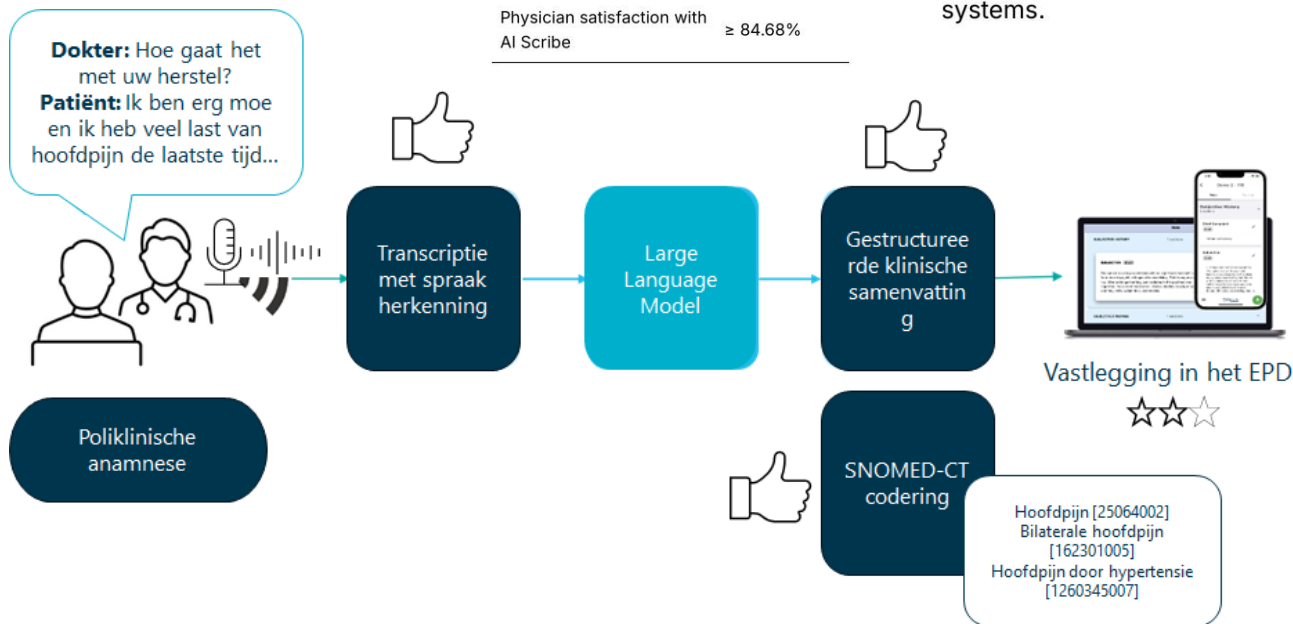
Validatie...?

5. Performance Characteristics

Clinical Performance Characteristics	Minimum Standard (Non-inferiority margin)
Reduction of average time in notes per appointment / patient with AI Scribe	≥ 0.57 min
Physician satisfaction with AI Scribe	$\geq 84.68\%$

Precautions

- AI may introduce transcription or summarisation errors.
- Do not copy or transfer unverified content into medical records or other downstream systems.





Stap 1. Bepaal de medische context



Stap 2. Bepaal het innovatieniveau



Stap 3. Bepaal de toegevoegde waarde



Stap 4. Bepaal de risico's & impact



Stap 5. Maak afspraken over taken, bevoegdheden & verantwoordelijkheden



Stap 6. Controleer toepasselijke wet- & regelgeving



Stap 7. Bereid voor op implementatie



Fase 2: Ingebruikname

Wanneer is de AI-toepassing veilig genoeg om in gebruik te nemen?



Stap 8. Evalueer de uitkomsten



Fase 4: Afvoer

Leidraad Software validatie als basis



Richtlijndatabase

[RICHTLIJNEN](#) [Q ZOEK](#) [NIEUWS](#) [INSTRUCTIES](#) [WERKWIJZE](#) [OVER](#) [CONTACT](#) [APP](#)

[LOG IN](#) [COVID-19](#)

[← Terug naar zoekresultaten](#)

Leidraad Validatie van Software als Medisch Hulpmiddel

Initiatief: NVKF

Aantal modules: 2

[+ VOLGEN](#)

[Download richtlijn](#)

Leidraad Validatie van Software als Medisch Hulpmiddel

Zoeken binnen deze richtlijn



1. Startpagina - leidraad validatie van software als medisch hulpmiddel

Startpagina - leidraad validatie van software als medisch hulpmiddel

Beoordeeld: 12-09-2024

Wat is nieuw?

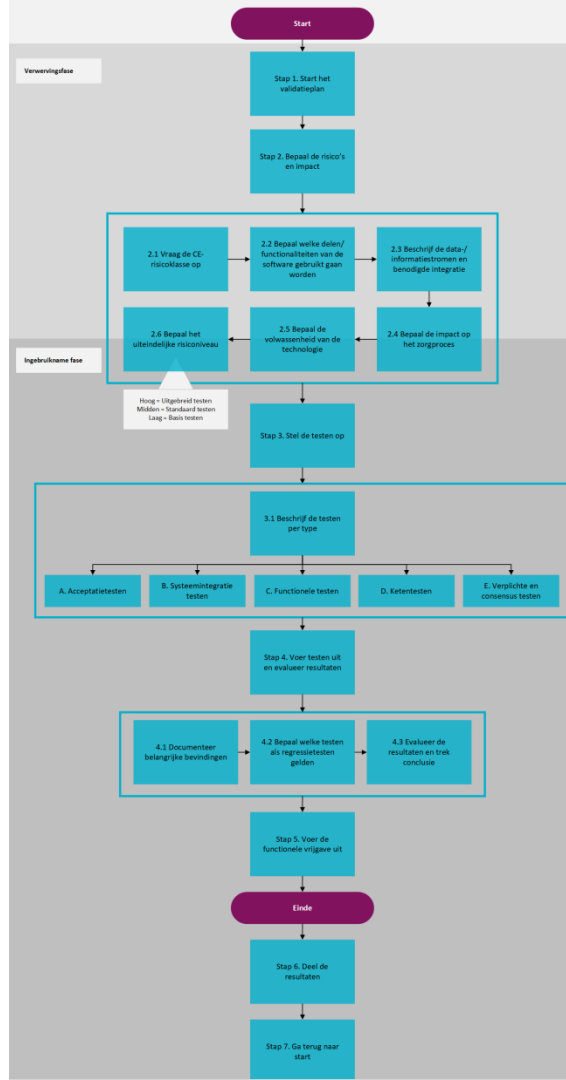
[Startpagina - leidraad validatie van software als medisch hulpmiddel](#)

[Leidraad](#)

Publicatiedatum

28-10-2024

28-10-2024



Lokale validatie bij AI extra belangrijk

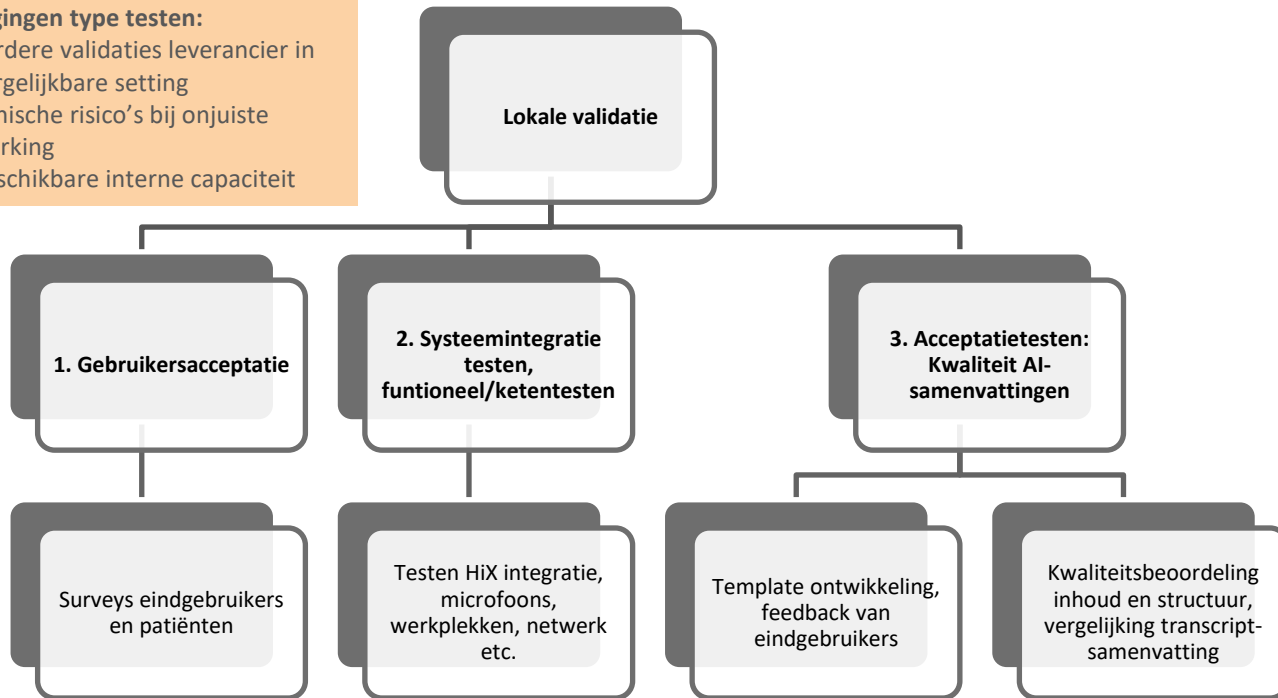
- Is vaak méér dan alleen testen van software en koppelingen
- Performance van AI-modellen op lokale data
- Frequenter her-validatie nodig bij updates en upgrades, modelveranderingen
- Meer vraagstukken omtrent dataveiligheid en privacy
- Nieuwe (generatieve) AI-technologie
 - Onvolwassenheid leveranciers en markt in ontwikkeling
 - In geval van LLM's methodologische uitdagingen



Onze aanpak...

Afwegingen type testen:

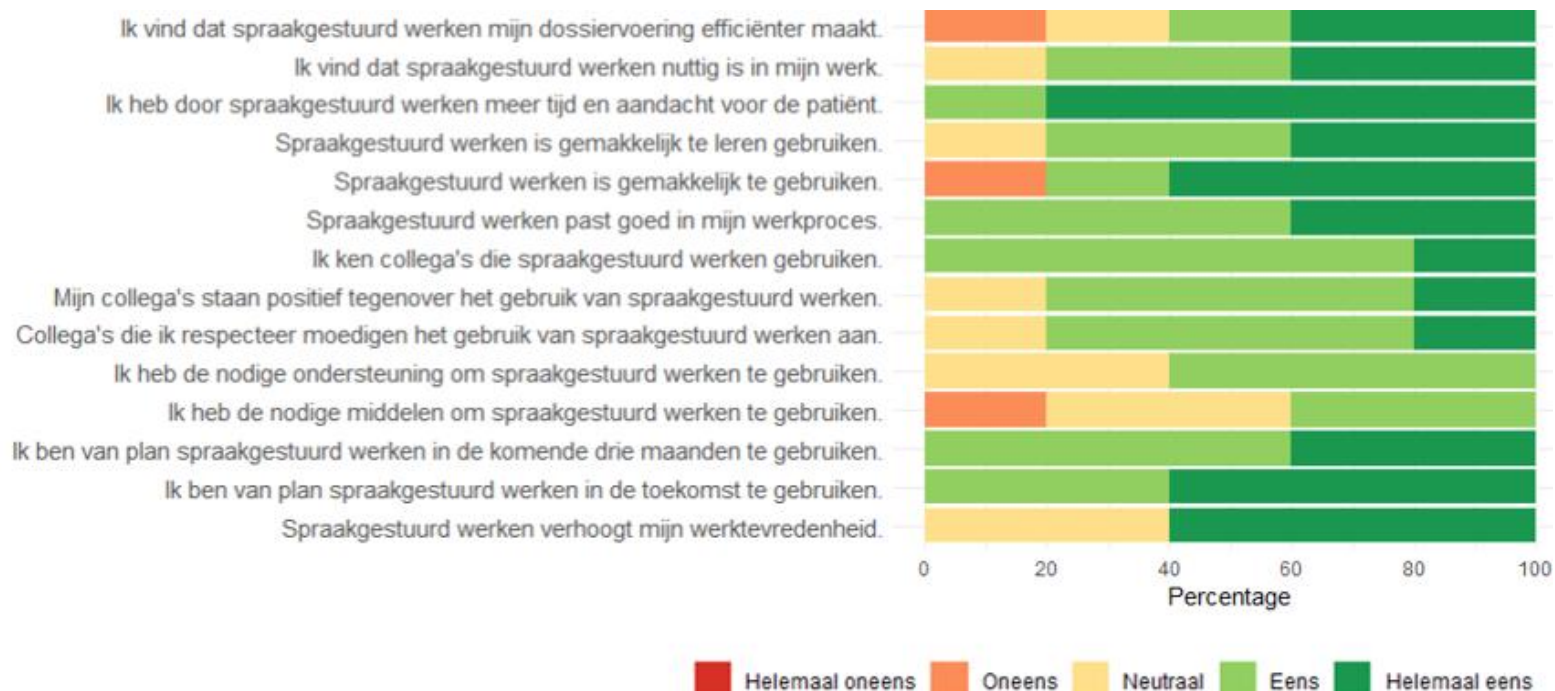
- Eerdere validaties leverancier in vergelijkbare setting
- Klinische risico's bij onjuiste werking
- Beschikbare interne capaciteit



Opzet validatie gebruikerservaring

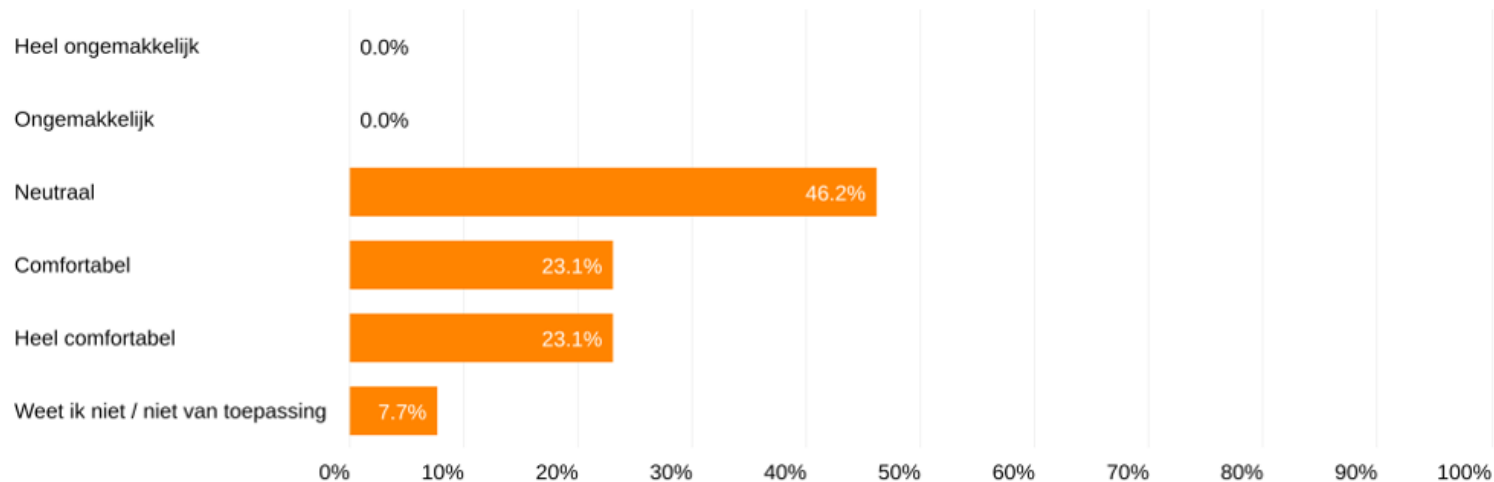
Groep		Doel	Instrument	Respondenten	Periode
Artsen	Gebruikers	Inventariseren ervaringen met gebruik SPW en houding t.o.v. van verdere uitrol binnen eigen vakgroep (gericht op minimaal product)	UTAUT NPS	5	7-17 november
			Semi-gestructureerde interviews	8	3-24 november
	Niet-gebruikers	Verkennen verwachtingen en intentie tot gebruik SPW	UTAUT	17	7-17 november
Patiënten		Evaluëren beleving consult met SPW	Vragenlijst	14	3-17 november

Gebruikers



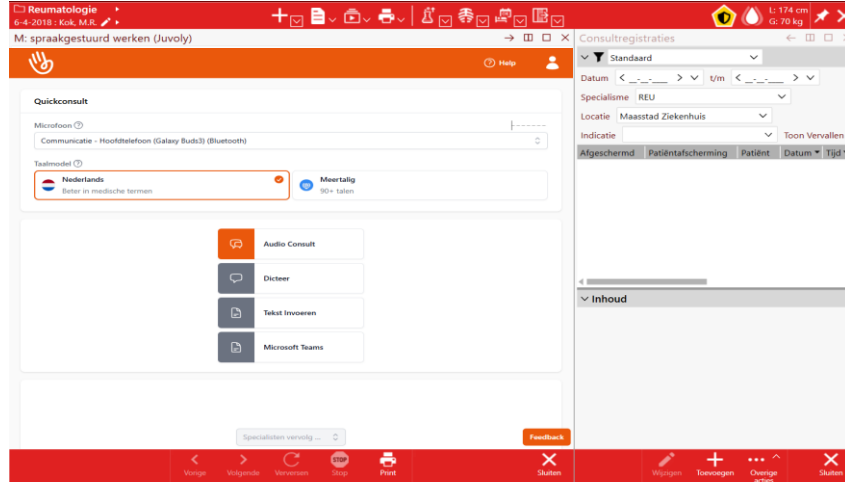
Patiënten

Hoe voelt u zich bij het gebruik van een computer die meeluistert en meeschrijft (AI)?

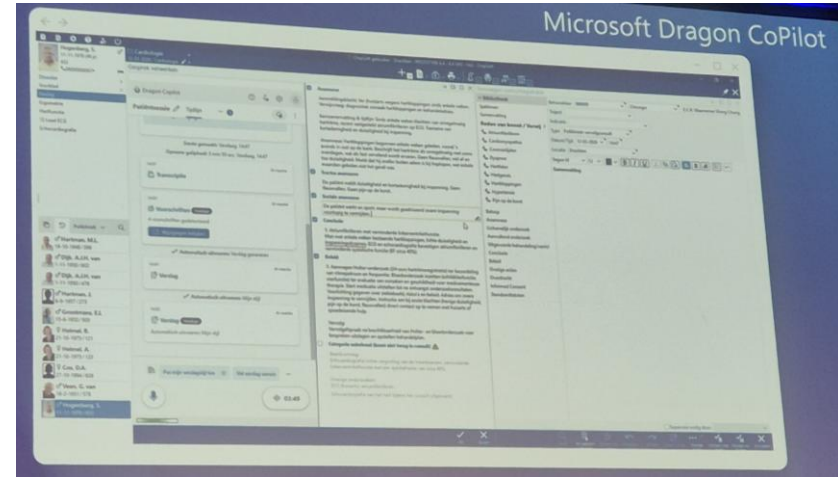


Functionele test: integratie in EPD

Huidige situatie (iFrame)



Toekomstige situatie



Waarom kwaliteit van AI-samenvattingen valideren?

Type fout	Voorbeeld arts	AI-samenvatting
Missende informatie	"Vervolgafspraak over 3 maanden"	-
Hallucinaties	- (dermatologie consult)	Tr. locomotorius: geen gewrichtsklachten. Tr. respiratorius: geen kortademigheid. ...
Ontbreken nuance	"Verzuimt bijna nooit maar het is echt lastig soms"	"Verzuimt bijna nooit vanwege klacht. Sociale activiteiten gaan door ondanks klacht."
Tegenstellingen, wel/geen	"3 Keer aura gehad: neonachtige vlek in beeld"	"Geen aura, soms neonachtige vlekken in beeld"
Familierelaties	"Moeder heeft migraine, vader van moeder had het"	"Vader migraine, zus had migraine"
Medicatie namen	"eletriptan"	"zomitriptan"
...		

Steeds meer bewijs...

70% van draft notes
bevatten fouten
(n = 44) onderzocht met 2
commerciële AI-tools

Missende informatie
(ommissies) komen
het vaakst voor



Adoptie valt tegen, 32% gebruikt de AI-tool bij de helft of meer consulten

JMIR Publications
Advancing Digital Health & Open Science

Articles ▾ Search articles

Journal of Medical Internet Research

Journal Information ▾ Browse Journal ▾

Published on 27-Jan-2025 in Vol 27 (2025)
Preprints (earlier versions) of this paper are available at <https://preprints.jmir.org/preprint/64993>, first published 01.Aug.2024.

Accuracy and Safety of AI-Enabled Scribe Technology: Instrument Validation Study

Joshua Biro¹, Jessica L Handley¹, Nathan K Cobb^{2,3}, Varsha Kottamasu², Jeffrey Collins², Seth Krevat^{1,3}, Raj M Ratwan^{1,3}

Article Authors Cited by (18) Tweetatons (2) Metrics

npj | digital medicine

Explore content ▾ About the journal ▾ Publish with us ▾

nature > npj digital medicine > comment > article

Comment | [Open access](#) | Published: 24 September 2025

Beyond human ears: navigating the uncharted risks of AI scribes in clinical practice

[Maxim Topaz](#) ✉, [Laura Maria Peltonen](#) & [Zhihong Zhang](#)

[npj Digital Medicine](#) 8, Article number: 569 (2025) | [Cite this article](#)

16k Accesses | 15 Citations | 54 Altmetric | [Metrics](#)

> J Contin Educ Nurs. 2025 Sep;56(9):358-359. doi: 10.3928/00220124-20250814-03. Epub 2025 Sep 1.

Invisible Scribes: Can Nurses Trust Ambient AI for Clinical Documentation?

Maxim Topaz ¹

Affiliations + expand

PMID: 40857680 DOI: 10.3928/00220124-20250814-03

AI beoordeelt AI (LLM as a judge)...











CURRENT ISSUE ▾ RECENTLY PUBLISHED

[DATASETS, BENCHMARKS, AND PROTOCOLS](#)



Verifying Facts in Patient Care Documents Generated by Large Language Models Using Electronic Health Records

Authors: Philip Chung, M.D., M.S. , Akshay Swaminathan , Alex J. Goodell, M.D., M.S. , Yeasul Kim, M.D., M.S. , S. Morsen Reincke, M.D. , Lichy Han, M.D., Ph.D. , Ben Devereett, M.D., Ph.D. ,  ⁺²¹, and Nima Aghaeepour, Ph.D.  [Author Info & Affiliations](#)

Published December 24, 2025 | NEJM AI 2026;3(1) | DOI: 10.1056/AIdbp2500418 | [VOL. 3 NO. 1](#)

[Copyright © 2025](#)



Abstract

BACKGROUND

Large language models (LLMs) can generate clinical documents, but inaccuracies due to hallucinations create a barrier for use in medical practice. Reliance on clinician vigilance to remedy errors is infeasible, since mistakes may be subtle and interleaved among accurate statements within the document. An increasing trend in artificial intelligence (AI) has been to deploy LLMs as evaluators of LLM output (LLM-as-a-Judge), which has been shown to have high concordance with human judgment.

METHODS

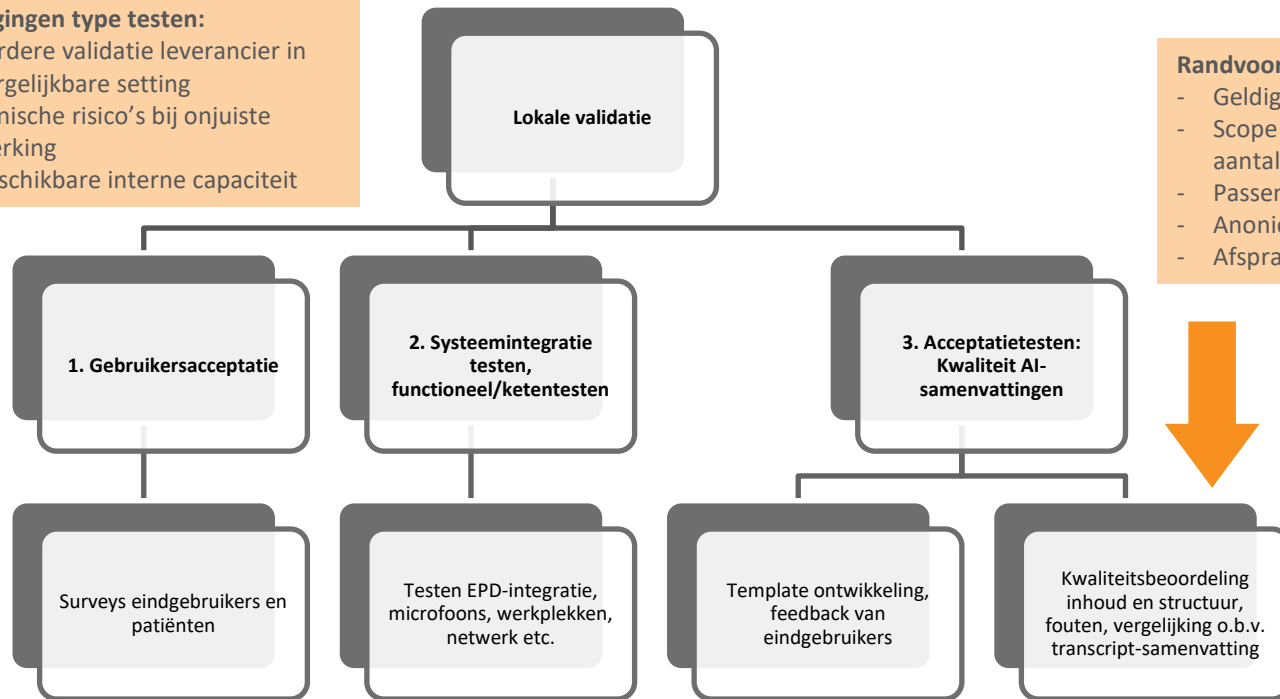
We hypothesize that an LLM-as-a-Judge with relevant facts from a patient's electronic



Onze aanpak...

Afwegingen type testen:

- Eerdere validatie leverancier in vergelijkbare setting
- Klinische risico's bij onjuiste werking
- Beschikbare interne capaciteit



Randvoorwaarden:

- Geldige verwerkersovereenkomst
- Scope bepalen (functionaliteiten, templates, aantallen)
- Passende grondslag AVG (toestemming)
- Anonieme transcripten en AI-samenvattingen
- Afspraken met leverancier over ondersteuning

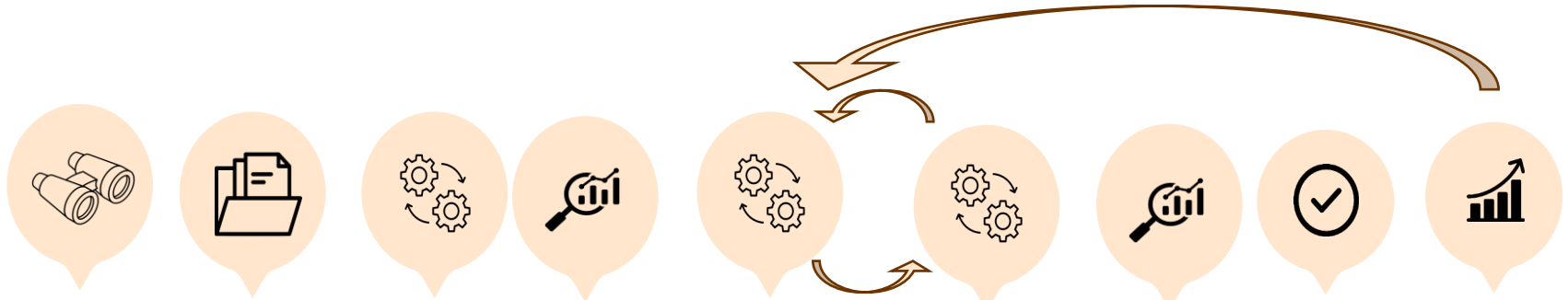


Minimaal bekend bij leverancier:

- Dataset (herkomst, aantal) en evaluatiemethode
- Kwaliteit transcriptie
- Kwaliteit inhoud & structuur AI-samenvattingen
- Hallucination rate (%)
- Omission rate (%)
- Clinically relevant errors (%)
- Bias detectie (%)

- Op welke manier aanbieden tandem?
- Aantallen patienten/transcripten?
- Methode: inhoud vs kwaliteit?

(concept) stappenplan validatie



Step 0	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Step 7	Step 8
inventarisatie	data verzamelen	template optimalisatie	validatie LLM	praktijk test	template optimalisatie	validatie LLM	acceptatie en vrijgave	wijzigingen?
Huidige werkwijze inzichtelijk maken	Verzamel en anonimiseer data van +/- 10-20 patienten:	Leverancier optimaliseert template obv data uit stap 1	Doel: vergelijk vorige versie	Doel: optimalisatie inhoud	Leverancier optimaliseert template obv data uit stap 1	Doel: veilig in gebruik nemen nieuwe versie	Template wordt niet meer gewijzigd	Hergebruik data & methodiek bij:
Wat zijn eisen/wensen?	- Transcript	Resources: Tandem	Methode: inhoud en kwaliteit	+/-10 artsen volledig over naar Tandem met template	Inhoud 1. samenvatting vergelijking EPD vs LLM	Methode inhoud (uit stap 5) en kwaliteit	Resources: MSZ	- Wijzigingen
Resources: Tandem+MSZ	- Samenvatting LLM		Resources: Tandem	-Feedback op template	Resources: MSZ+Tandem	Data: nieuwe testset volgens stap 1		- Nieuwe templates
	- Beoordeling medisch specialist op samenvatting LLM			Resources: MSZ		Resources: MSZ+Tandem		- Etc..
	- Samenvatting in EPD							Resources: MSZ+Tandem
	Resources: Tandem							

Leren door te doen!

Data en aanpak:

- Aantal consulten: PEC (7x), PVC (6x)
- Beoordeling door medisch specialist (0=slechte score - 2=prima score)
- Weging inhoud (70%), kwaliteit (30%)

Casus ID	Consult Juvoly	Consult Tandem	Toelichting
1. Inhoud	Score (0-2)	Score (0-2)	
Reden van komst	1	2	
Anamnese (incl tractus/familie)	1	2	
Sociale anamnese	2	2	
Medicatie	nvt	nvt	
Voorgeschiedenis	0	1	
Intoxicaties	nvt	nvt	
Allergieën	nvt	nvt	
Lichamelijk onderzoek	2	2	
Aanvullend onderzoek	2	2	
Conclusie	1	2	
Differentiaal diagnose	1	2	
Diagnose	1	2	
Beleid	2	1	
Informed consent	1	2	
Totaal=	14	20	
2. Kwaliteit	Score (0-2)	Score (0-2)	
Structuur	1	2	Zinsopbouw, 2p = soepel/goed lopend verhaal
Beknoptheid	2	1	2p = Niet te lang, niet te kort
Geen informatie dubbelop	2	1	2p = geen dubbele info, 1p = 1 dubbel informatie, 0 = >1 dubbele informatie
Geen hallucinaties	2	2	2p = geen hallucinaties, 1p = 1 hallucinatie, 0 = meer dan 1 hallucinatie
Grammatica/spelling	1	2	2p = helemaal goed, 1p = (paar) kleine fouten, 0 = slechte grammatica/spelling
Totaal =	8	8	

Berekening = aantal items die voorkomen in het consult maken samen de totale aantal punten die behaald kunnen worden.
Daarna aantal punten gescoord/totaal aantal punten x100%. Weging van inhoud is 70% en kwaliteit 30%

	Consult Juvoly	Consult Tandem
Inhoud	14/22x100%=64%	20/22x100%=91%
Kwaliteit	8/10x100%=80%	8/10x100%=80%
Totaal=	(0.7x64)+(0.3x80)=69%	(0.7x91)+(0.3x80)=88%

	Inhoud Juvoly	Kwaliteit Juvoly	Totaal Juvoly	Inhoud Tandem	Kwaliteit Tande	Totaal Tandem
Casus pericarditis	76%	60%	76%	71%	50%	65%
Casus Coeliakie	41%	90%	56%	82%	50%	72%
Casus Frozen shoude	80%	60%	74%	80%	79%	77%
Casus TIA pec	65%	80%	70%	85%	50%	75%
Casus BPPD pec	71%	70%	71%	75%	60%	71%
Casus syndroom van c	65%	70%	67%	75%	50%	68%
Casus Basaalcelcarcin	65%	80%	70%	80%	90%	83%
Gemiddeld	66%	73%	69%	78%	61%	73%

Maasstad Apotheek

Dit doen we niet alleen...

[RIGH:T consortium](https://www.Rightconsortium.nl)

www.Rightconsortium.nl



RIGH:T consortium

Het Nederlands Adaptief Consortium voor Verantwoorde Implementatie van Generatieve AI in de Zorg (Engels: *Responsible Implementation of Generative AI in Healthcare: Together*, oftewel *RIGH:T*) is een samenwerkingsverband tussen zorginstellingen en stakeholders ter bevordering van de verantwoorde implementatie van generatieve *artificial intelligence* (AI), en in het bijzonder grote taalmodellen (Large Language Models, LLMs) in de zorg.

Het consortium wil de gehele lifecycle (na ontwikkeling) van LLMs in zorg waarborgen.



Wat doet het
RIGH:T consortium

Projecten



Participerende centra



Elisabeth-TweeSteden Ziekenhuis



Erasmus MC



UMC Groningen



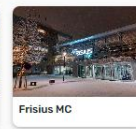
GGZ Noord-Holland Noord



Levvel



TU Delft



Frisius MC



Leiden UMC



UMC Utrecht



Maastricht Ziekenhuis

Aangesloten stakeholders

LHV

NVZ

NFU

VWS

CMIO netwerk GGZ

VVT

Nivel

Patiëntenfederatie

TNO

NHG

Digizo

Take home message

- Generatieve AI wordt **al toegepast in de klinische praktijk**
- **Lokale validatie** bij AI extra belangrijk
- De markt, de technologie en de (performance) validatie methodiek zijn **nog in ontwikkeling**
 - Zoek samenwerking met AI-leverancier en bespreek verantwoordelijkheden
 - Validatie is bij deze technologie niet eenmalig: doe dit voor ingebruikname, bij wijzigingen en tijdens gebruiksfase door kwaliteit te monitoren
- Bij medical device is validatie een vereiste: **stappenplan healthy AI** en **leidraad software validatie** zijn handige hulpmiddelen
- Jullie expertise, kennis en ervaring bij testen en vrijgave van medisch technologie helpt bij deze nieuwe technologie, zeker als dit als medical device software wordt ingezet

MAASSTAD ZIEKENHUIS

een santeon ziekenhuis



Lokale validatie van medische AI

René Verhaart en Sade Krijgsman

klinisch informatici

25 februari 2026



MAASSTAD ZIEKENHUIS

een santeon ziekenhuis



werkenbijmaasstadziekenhuis.nl

Boven alles
Rotterdams