



Whitepaper

November 2025

AI in de Spreekkamer

Inhoud

1. Voorwoord	3
2. Inleiding	4
3. Werkwijze	5
3.1 Innovation Call	
3.2 mProve transformatieproject 'Transcriptie in de Spreekkamer	
3.3 AI Validatielab-sessie	
3.4 Beoordeling en vertrouwelijkheid	
4. Marktontwikkelingen	7
4.1 Ontwikkelingsfase	
4.2 Impact	
4.3 Integratie in HiX	
4.4 NEN 7510 Certificering	
4.5 Medical Device Regulation (MDR)	
4.6 Taalmodel	
4.7 Doorontwikkeling	
4.8 Supportorganisatie	
4.9 Prijsmodel	
5. Meetresultaten	14
5.1 Snelheid	
5.2 Toetsing casus Longgeneeskunde	
5.3 Toetsing verpleegkundige casus	
6. Takeaways	22

1. Voorwoord

Beste lezer,

Voor u ligt de tweede whitepaper 'AI in de Spreekkamer' vanuit het mProve Funxlab. De eerste whitepaper van maart 2025 bood ons waardevolle inzichten over de kracht van AI om administratieve lasten te verlichten, maar ook over de grenzen van de huidige technologie. We leerden dat kwaliteit, betrouwbaarheid en integratie geen vanzelfsprekendheden zijn, maar uitdagingen die we samen moeten aangaan. En juist dat 'samen' is waar mProve sterk in is: kennis delen, samen leren, en de lat telkens een stukje hoger leggen.

In dit vervolg bouwen we voort op onze eerste whitepaper. We verkennen hoe de toekomst eruitziet als AI niet langer een experiment is, maar een betrouwbare virtuele collega die onze zorgprofessionals ondersteunt bij administratieve taken.

Namens de 7 mProve ziekenhuizen en partners dank ik iedereen die deze innovatiereis mede mogelijk maakt. De toekomst van zorginnovatie is geen kwestie van techniek alleen, het is een kwestie van durf, vertrouwen en samenwerking. En daar leveren we als mProve ziekenhuizen graag een impactvolle bijdrage aan!

Stefan Kroese

Voorzitter mProve/ voorzitter raad van bestuur Noordwest Ziekenhuisgroep

2. Inleiding

De spreekkamer is één van de vitale plekken van het ziekenhuis. Hier worden diagnoses gesteld, behandelplannen besproken en beslissingen genomen met grote impact op het leven van patiënten. Hoe kan kunstmatige intelligentie (AI) daarbij helpen? De belofte is groot: AI kan artsen ondersteunen bij diagnostiek, patronen herkennen die mensen ontgaan en administratieve taken verlichten. In theorie betekent dit meer tijd voor persoonlijke zorg en betere besluitvorming.

Maar hoe ver zijn we in de praktijk daarmee? De markt voor AI-oplossingen groeit razendsnel en steeds meer leveranciers beloven de zorg te kunnen transformeren.

Vanuit het mProve Funxlab organiseerden we eind 2024 een Innovation Call. Met deze oproep aan het bedrijfsleven riepen we leveranciers op om hun toepassing tijdens een live AI Validatielab-sessie te demonstreren. Dit gebeurde in februari 2025 in een unieke en vertrouwelijke co-creatie setting waarbij mProve inzicht en kennis kreeg over de stand van zaken in de markt. Leveranciers kregen individueel concrete (door-) ontwikkelsuggesties voor hun product.

In oktober 2025 hebben we deze sessie nog eens dunnetjes overgedaan, zodat we trends en ontwikkelingen goed kunnen volgen. In deze whitepaper vind je onze bevindingen, conclusies en aanbevelingen. Deze delen wij graag met de Nederlandse zorg om zo bij te dragen aan de (door-) ontwikkeling en opschaling van deze impactvolle technologie. We gaan in op hoe de markt er op dit moment voor staat, geven inzicht in de prestaties van spraakgestuurde AI-toepassingen voor de spreekkamer (en daarbuiten) en doen suggesties voor de richting waarop wij zien dat deze technologie impactvol kan zijn.

We danken alle collega's uit de mProve ziekenhuizen die hebben meegewerkt aan deze sessie en dit rapport. Daarnaast danken we uiteraard ook de dappere ondernemers van *Autoscriber*, *OurMind*, *Tandem Health*, *Noteless*, *Valuecare* en *G2Speech* voor hun deelname, en vijf collega-startups voor hun aanmelding. Het starten van een technologiebedrijf in de Nederlandse zorg vraagt visie, moed en een ongelofelijke dosis doorzettingsvermogen. Het live demonstreren van die technologie met alle afbreukrisico's van dien nog nét iets meer.

Jos Knaapen

Programmamanager mProve Funxlab

3. Werkwijze

3.1 Innovation Call

Via een 'Innovation Call' riep mProve voor de tweede keer start- en scale-ups op om hun product voor 'AI in de spreekkamer' tijdens een live AI Validatielab-sessie te laten beoordelen. Dit keer reageerden 11 leveranciers op deze oproep. Zij vulden een intake-vragenlijst in met daarin onder andere relevante informatie over de manier waarop hun product of dienst werkt, hoe hun prijsmodel in elkaar zit, over hun visie op integratie in het EPD en hun ideeën voor *use cases* voor hun technologie buiten de poliklinische setting. Uit die aanmeldingen zijn 6 leveranciers, met een bewezen track record in de medisch specialistische zorg in Nederland of daarbuiten, uitgenodigd voor de sessie van 6 oktober 2025 in het Jeroen Bosch Ziekenhuis.

Tijdens deze sessie werden presentaties gegeven door het Jeroen Bosch Ziekenhuis en Noordwest Ziekenhuisgroep over onder andere de gemeten impact van deze technologie. Leveranciers pitchten volgens een strak format met verplichte onderdelen en stelden hun meest prangende vraag aan het publiek. Daarnaast voerden we onze periodieke meting van de kwaliteit van de technologie uit door twee gesprekken te verslaan. De daaruit voortvloeiende verslagen zijn beoordeeld op compleetheid, correctheid, structuur en formulering. Daarnaast hebben we onze beoordelaars, allen zorgverlener in een mProve ziekenhuis, gevraagd om een inschatting van tijdsbesparing te maken. We zijn bij deze sessie ook weer buiten de spreekkamer gegaan voor een verpleegkundig experiment.

3.2 mProve transformatieproject 'Transcriptie in de spreekkamer'

Het project 'Transcriptie in de spreekkamer' maakt onderdeel uit van het mProve Transformatieprogramma, dat in maart 2025 werd goedgekeurd door de betrokken zorgverzekeraars. Bij onze eerste 'AI Validatielab'-sessie van februari 2025 waren 4 van de 7 mProve ziekenhuizen bezig met een pilot, POC, evaluatie of opschaling van deze technologie, met een tweetal verschillende leveranciers. Dat is in lijn met de manier waarop we binnen mProve werken: 2 ziekenhuizen starten met een innovatieproject, en stapsgewijs volgen de anderen bij succesvolle resultaten uit pilotperioden. In februari waren de eerste twee '*volgende*' ziekenhuizen dus gestart.

In de zomer hebben we samen geconcludeerd dat deze technologie impactvol genoeg is om in de andere mProve ziekenhuizen ook een GO te geven op het starten van hun implementatie. Inmiddels zijn 6 van de 7 ziekenhuizen live met de technologie van *Autoscriber* en *Ourmind*, beiden deelnemer aan onze AI Validatielab-sessie. Het 7^e mProve ziekenhuis volgt dit jaar nog met hun livegang met deze technologie.

We komen in het project tot de conclusie dat deze technologie leidt tot een aantoonbare verlaging van de werkdruk en een verhoging van het werkplezier. Wel is aanvullend onderzoek met een grotere groep respondenten nodig om onze (wetenschappelijk significante) resultaten verder te onderbouwen. We zien dat deze technologie het beste ingezet wordt bij consulten met nieuwe patiënten, het liefst in specialismen met gemiddeld een langere consultduur. Daarnaast zien we steeds meer impact bij verpleegkundigen, verpleegkundig specialisten en paramedici. Ook zien we dat de eerste ervaringen met de toepassing door verpleegkundigen hebben geleid tot onder meer een finaleplaats voor de Noordwest Ziekenhuisgroep in de STZ Innovatiechallenge; de innovatieprijs van de 27 samenwerkende topklinische ziekenhuizen.

3.3 AI Validatielab-sessie

De leveranciers lieten tijdens de sessie hun software meedraaien tijdens twee gesprekken van een patiënt en een zorgverlener, aangevuld met een of meerdere mantelzorgers. Tijdens deze sessie hebben we het specialisme Longgeneeskunde en een 'Eerste verantwoordelijke verpleegkundige'-gesprek (een EVV-gesprek) uitgelicht. Het EVV-gesprek is door het vaste aanspreekpunt in de zorg om het zorgplan met een patiënt te bespreken, evalueren en aan te passen.

Na afronding van de gesprekken werd de samenvatting zo snel mogelijk gedeeld via een daarvoor bestemd e-mailadres. Daarmee konden we ook de doorlooptijd tot resultaat van de AI-modellen toetsen. In de praktijk moet de technologie immers direct goed functioneren; modellen die pas na langere tijd een verslag opleveren, zijn in de zorg beperkt bruikbaar. Zeker in de spreekkamer, waar de omloopsnelheid van patiënten hoog is, is snelheid cruciaal.

3.4 Beoordeling en vertrouwelijkheid

Achteraf beoordeelden een aantal artsen en verpleegkundigen uit de mProve ziekenhuizen de compleetheid, correctheid, structuur, formulering en algehele bruikbaarheid van de samenvattingen. En ook dit keer gingen we verder: data scientists beoordeelden de achterliggende technologie en bedrijfskundigen en ICT-collega's het bedrijfsmodel, supportlevel, de betrouwbaarheid en compliance. Leveranciers ontvingen van ons een speciaal voor hen opgesteld (vertrouwelijk) rapport met daarin waardevolle doorontwikkelsuggesties voor hun product.

Een labsetting als deze kan alleen maar op basis van vertrouwelijkheid. Er is nog steeds veel hype rondom deze technologie en er zijn hoge verwachtingen voor de toekomst (en in het heden), maar de technologie is voorlopig nog niet af. Alleen door samen te werken kan deze technologie de volgende kwaliteitssprong maken en daarmee ook een grote impactsprong. Daar past een vertrouwelijke setting bij die is gebaseerd op wederzijdse ambitie om de zorg te verbeteren. De resultaten in dit rapport zijn daarom niet direct terug te leiden naar individuele leveranciers.

4. Marktentwikkelingen

Onze algemene bevindingen zijn verzameld via een zelfevaluatie door de betrokken leveranciers. Aan de hand hiervan schetsen we de status van de markt.

4.1 Ontwikkelingsfase

We zien dat de leveranciers ten opzichte van onze voorgaande meting vaker de status van opschaling bereiken. Zij hebben allemaal de 'ontdekkingfase' verlaten en zijn in de validatiefase of verder uitgekomen. Dit herkennen we in de praktijk ook in het mProve project 'Transcriptie in de Spreekkamer', waar nagenoeg alle mProve ziekenhuizen inmiddels gebruik maken van de technologie.

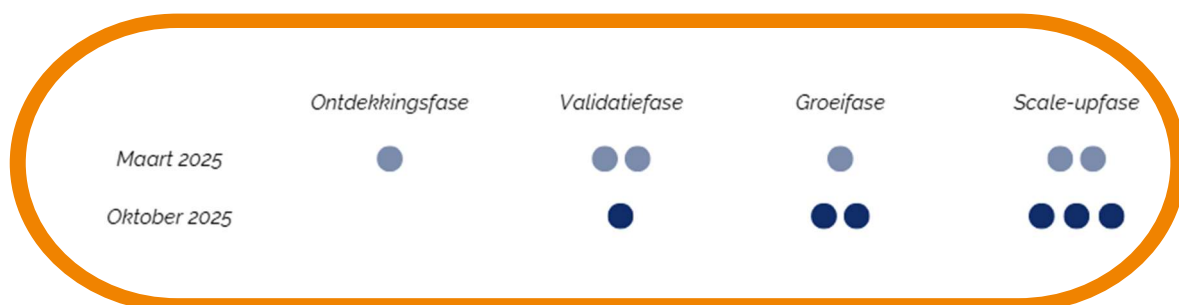
Drie van de zes leveranciers beschouwt zichzelf als 'marktleider' in de medisch specialistische zorg in Nederland. In de praktijk zien we dat er nog stappen zijn rondom het aantal medisch specialisten en verpleegkundigen die de technologie op structurele basis in zijn of haar bedrijfsvoering heeft opgenomen.

We zien ook een nieuwe trend. Waar we in de voorgaande AI Validatielab-sessie ook aanmeldingen ontvingen van bedrijven die vanuit bijvoorbeeld huisartsgeneeskunde of GGZ-instellingen een zijdelingse stap beoogden naar de medisch specialistische zorg, lijken zij nu weer vooral op hun primaire sector te focussen. Daarentegen zien we nu internationale scale-ups die de Nederlandse markt benaderen. Daarbij maken onder andere scale-ups uit Scandinavië of het Verenigd Koninkrijk de overstap. Daarnaast zien we multinationals als Microsoft de stap naar Nederland maken.

Sommige leveranciers hebben in het buitenland investeringen van tientallen miljoenen opgehaald voor internationale expansie. Dat zijn investeringen die we in de Nederlandse context minder gewend zijn. We gaan de komende maanden tot jaren zien welke impact dat op de marktentwikkeling in Nederland gaat hebben.

Conclusie

De markt is volop in beweging. In de afgelopen maanden zien we meer internationale activiteit uit andere Europese lidstaten of vanuit de Verenigde Staten. We zien tegelijkertijd dat de Nederlandse ziekenhuizen zelf ook stappen zetten van pilotfase richting opschaling. De markt wordt op dat gebied langzaam volwassen. Tegelijkertijd horen we van ziekenhuizen dat ze, gezien de snelle ontwikkeling van de markt, relatief korte contracten afsluiten met leveranciers. We verwachten door internationale ontwikkelingen dat de markt voor deze technologie voorlopig nog wel in beweging blijft.



Figuur 1: Ontwikkelingsfase product

4.2 Impact

Waar we in maart nog niet veel (wetenschappelijk gevalideerde) metingen terugzagen, zien we nu steeds meer bewijslast voor de inzet van deze technologie. Zo heeft Noordwest Ziekenhuisgroep een 'stopwatchmeting' gedaan, waarbij de impact van de technologie nauwgezet is gemeten. De mProve ziekenhuizen hebben inmiddels één uniforme manier van nul- en effectmeting opgezet. We verwachten dat de vervolgmetingen betere resultaten laten zien als de technologie geïntegreerd in het elektronisch patiëntendossier (EPD) werkt. Op dit moment vinden er bij verschillende ziekenhuizen nog handmatige copy-paste handelingen plaats om verslagen in het EPD te krijgen.

Zonder de onderzoeken in één keer plat te willen slaan zien we dat deze technologie idealiter wordt ingezet bij *nieuwe patiënten* bij *beschouwende specialismen*. We zien een causaal verband tussen de lengte van een consult en de hoeveelheid nieuwe informatie die een patiënt deelt met de zorgverlener, en die in een verslag verwerkt moet worden. Dat is een open deur, maar tegelijkertijd zien we behoorlijk wat zorginstellingen die de technologie in een pilotperiode inzetten bij bijvoorbeeld herhaalconsulten van een snijdend specialisme. Daarnaast horen we uit ons gezamenlijke mProve project positieve geluiden terug van verpleegkundig specialisten en medisch psychologen, die de technologie steeds meer omarmen vanwege de impact op registratietijd, werkplezier en -druk.

Verschillende aanbieders geven aan dat hun oplossingen bijdragen aan een reductie van administratietijd, variërend van enkele minuten per consult tot een besparing van 25 tot 30% van de administratietijd. Wij zien deze cijfers niet altijd even goed onderbouwd terug in de onderzoeken die zijn gedaan in Nederland of daarbuiten. Tegelijkertijd constateren we dat onderzoeksresultaten die nu getoond worden al verouderd zijn bij publicatie.

Een deel van de leveranciers leverde geen impactmeting aan bij het aanmelden voor onze Innovation Call. Alleen actiegericht onderzoek met korte doorlooptijd leidt tot de bouwstenen die ziekenhuizen nu zoeken: de ingrediënten die leiden tot een inschatting van de (structurele) business case van deze technologie.

Conclusie

Niet alle leveranciers kunnen gemeten effectcijfers van hun technologie aantonen. In lijn met onze vorige whitepaper adviseren wij leveranciers om de beoogde impact, in samenwerking met zorg- en onderzoeksinstituten, verder te valideren.

4.3 Integratie in HiX

Zoals eerder geconcludeerd: de effectiviteit van de software neemt aanzienlijk toe zodra deze volledig is geïntegreerd in het elektronisch patiëntendossier. Hierdoor hoeven zorgverleners niet langer handmatig samenvattingen te kopiëren en plakken in het medisch dossier.

Twee mProve ziekenhuizen hebben inmiddels een integratie van hun spraakoplossing in HiX. De rest van de ziekenhuizen wachten op de integratie van de technologie via een API-koppeling, aangezien dat de enige koppeling is die in de toekomst wordt aangeboden. Op die manier heeft echter nog geen enkele leverancier of ziekenhuis de technologie geïntegreerd in HiX. Sinds onze sessie in maart zien we – en dat vinden we

zeer spijtig - geen schaa sprong in de integratie van deze technologie in het EPD. Daarbij kijken we eerst over de integratie van vrije tekstvelden via het consultsjabloon in HiX. We zijn nog niet aangekomen bij verdergaande integraties met gestructureerde datapunten, of zelfs inclusief het klaarzetten van orders – iets wat op verschillende roadmaps van leveranciers is opgenomen. De start- en scale-ups ontwikkelen significant sneller dan de integratiesnelheid in de praktijk mogelijk maakt.

Als we om ons heen kijken wijken we in Nederland behoorlijk af van andere Europese landen: de gemiddelde tijd voor integratie van de producten van de aanwezige buitenlandse leveranciers bedraagt over het algemeen gemiddeld een week. Dat zijn snelheden die we in Nederland niet bereiken.

Nagenoeg alle leveranciers in onze sessie zeggen te werken aan de integratie van hun product of dienst in HiX, via een API-koppeling via de DDA-module van Chipsoft. Dat deze al langer wordt aangekondigd maar nog steeds niet beschikbaar is, noch dat er duidelijkheid is wanneer leveranciers kunnen koppelen en ziekenhuizen daarmee gebruik kunnen maken van een geïntegreerde versie van de technologie, is op zijn minst bijzonder te noemen. De afhankelijkheid van ziekenhuizen van één of twee EPD-leveranciers die het tempo bepalen voor de zorg baart ons daarbij zorgen.

Conclusie

Het integreren in het EPD is de volgende grote stap voorwaarts voor alle leveranciers en gebruikers. Het handmatig knippen en plakken van transcripties of samenvattingen in het EPD is natuurlijk niet wat we willen in de zorg; integratie is van vitaal belang.

Het achterblijven van integratie zorgt ervoor dat we in Nederland nog lastig kunnen opschalen met deze technologie en onvoldoende betrouwbare impactmetingen kunnen doen van deze technologie. Leveranciers bouwen nieuwe impactvolle ontwikkelingen in hun standalone versie, zonder direct uitzicht op integratie daarvan in het EPD. Er is momentum voor deze technologie bij zorgverleners, die door beperkte integraties niet moet overslaan naar desillusie.

4.4 NEN7510 certificering

De NEN 7510-norm stelt eisen aan informatiebeveiliging in de zorgsector en waarborgt de vertrouwelijkheid, integriteit en beschikbaarheid van patiëntgegevens. Het behalen van deze certificering wordt steeds meer als randvoorwaarde gezien om de betreffende technologie te gebruiken in de Nederlandse zorg.

Gelijk aan de vorige keer in ons AI Validatielab zijn inmiddels 4 van de 6 partijen NEN 7510-gecertificeerd. Dit keer zijn er geen leveranciers die niet van plan zijn om de certificering aan te vragen. Ook de andere twee leveranciers verwachten binnen 1 jaar NEN 7510 gecertificeerd te zijn. Daarnaast hebben sommige aanbieders aanvullende certificeringen, zoals ISO 27001 en werken zij aan verdere kwaliteitsborging via ISO 13485.

Conclusie

Het voldoen aan de norm is van belang voor zorginstellingen, die nagenoeg alleen maar met gevoelige en vertrouwelijke medische informatie werken. We zien dat er gehoor is gegeven aan onze vraag om deze certificering serieus te nemen: het is een minimumvereiste aan het worden om 'mee te doen'. Want hoeveel impact een leverancier ook verwacht: de transcriptie moet op een veilige manier gebeuren.



Figuur 2: NEN 7510-certificering

4.5 Medical Device Regulation (MDR)

We hebben alle leveranciers gevraagd om hun visie op hun product in relatie tot de MDR te delen. En dat is een belangrijke vraag: we zien bij nagenoeg elke leverancier dat de volgende stap van hun technologie verder gaat dan alleen transcriberen en samenvatten. Het wegschrijven van orders in het elektronisch patiëntendossier, ondersteuning bij diagnosestelling en registratie worden door leveranciers genoemd als ontwikkelrichting. Daarbij bewegen we minimaal naar een grijs MDR-gebied. De technologie komt steeds meer in de buurt van wat als een medical device gezien wordt en is daarmee ook onderhevig aan de MDR.

We zien in de beantwoording op onze vraag dat leveranciers een andere afslag nemen als het om de MDR gaat in huidige of toekomstige functionaliteiten. Een deel van de leveranciers vindt het huidige product al een MDR Class 1-oplossing en heeft daarvoor een certificering (aangevraagd). Een ander deel concludeert dat de MDR niet van toepassing is op deze technologie.

Het grootste deel van de leveranciers werkt aan een 'AI Agent' die meer functies heeft dan de onderliggende technologie. En dat is natuurlijk fantastisch: hoe meer deze technologie als échte administratieve ondersteuner kan fungeren, hoe groter de impact ervan en hoe meer de technologie gewenst is.

Conclusie

Voor zover nog niet van toepassing roepen we leveranciers op om hun ontwikkelroadmap naast de MDR-vereisten te leggen. We constateren dat de pitches van leveranciers en bijbehorende roadmaps niet altijd in lijn zijn met de (huidige) visie van het bedrijf op de MDR. Dat beeld vullen we natuurlijk graag aan: een MDR-traject is complex, in de meeste roadmaps noodzakelijk, en kost veel tijd en geld.

4.6 Taalmodel

Om inzicht te krijgen in de achterliggende techniek van spraakgestuurde Gen AI-oplossingen is aan de leveranciers gevraagd naar de achtergrond en opbouw van hun Large Language Model (LLM). Deze is door een aantal data scientists uit verschillende mProve ziekenhuizen beoordeeld. We complimenteren de leveranciers daarbij graag met hun openheid en compleetheid in de beantwoording: we konden het, in vergelijking met voorgaande whitepaper, veel beter beoordelen.

De variatie in de backend van deze technologie is aanzienlijk. We zien nu behoorlijke verschillen in aanpak en wellicht ook bijbehorende kwetsbaarheid van de gekozen oplossing. We zien leveranciers die één bestaand taalmodel 'out of the box' gebruiken en hallucinaties onder controle houden door prompting - of volledig vertrouwen op de controlemechanismen van de leverancier van het betreffende taalmodel. Anderen gebruiken 'ensemble' methoden om verschillende taalmodellen te combineren voor een betere uitkomst en hebben een structurele automatische hallucinatiedetectie ontwikkeld. Deze laatste werkwijze is complexer dan de aanpak die bij andere leveranciers soms gehanteerd wordt. Er zijn in Nederland ook leveranciers die een eigen Nederlands taalmodel hebben ontwikkeld en gebruiken als achterliggend model. Helaas hebben die leveranciers zich niet aangemeld voor deze sessie.

Er is één duidelijke overeenkomst tussen de aanwezige partijen: de data wordt (volgens de zelfevaluatie) voor maximaal 24 uur bewaard en dan verwijderd. Het model wordt daarbij niet getraind op de data van klanten. Daarnaast gebruiken alle leveranciers in scope een Europees datacenter. Het gaat bij bijna alle leveranciers om het Europese datacenter van verschillende Amerikaanse techgiganten.

We zien dat er in Nederland, onder andere in de politiek, steeds meer vragen worden gesteld over de afhankelijkheid van Nederland van Amerikaanse techbedrijven. Bij deze technologie zien we die afhankelijkheid goed terug, evenals de worsteling om een goed Europees alternatief te vinden. Eén onderdeel waar Europa wél in voorop loopt¹ is regulering van AI. Om ook op andere fronten koploper te worden zien we dat Europa 'Apply AI Strategy' aan het ontwikkelen is, waarbij Europese startups gericht op de gezondheidszorg een belangrijke rol spelen.

Conclusie

Er worden door leveranciers verschillende afslagen genomen in de backend van de technologie waar een verschil in volwassenheid lijkt te bestaan. Daarbij zien we dat we in Nederland (of zelfs in Europa) afhankelijk zijn van Amerikaanse techreuzen. Tegelijkertijd constateren we dat er op dit moment weinig alternatieven beschikbaar zijn – en de roep om die alternatieven nog niet erg groot is.

4.7 Doorontwikkeling

De betrokken leveranciers hebben ook hun roadmap voor de toekomst gedeeld. De ontwikkelingen die zij als kansrijk achten, en waar in veel gevallen al actief aan gewerkt wordt, zijn uit te splitsen in de volgende hoofdlijnen:

Spraakgestuurde interactie

De technologie ontwikkelt zich van passieve transcriptie naar actieve spraakbesturing. Met gesproken opdrachten kunnen zorgverleners via spraak tekst bewerken of dicteren in elke applicatie. Dit maakt de AI breder inzetbaar bij de dagelijkse werkzaamheden en ondersteunt een volledige handsfree werkwijze, ook buiten het EPD. Dit sluit aan bij het beeld van 'ambient technologie', die op de achtergrond meewerkt.

¹ Bron: <https://nos.nl/nieuwsuur/artikel/2585550-europa-ontwaakt-in-de-ai-race-maar-druppel-op-gloeiende-plaat>

Automatisering en workflow-ondersteuning

Een belangrijk ontwikkelpad is het automatiseren van herhalende administratieve taken: het opstellen van verwijs- en ontslagbrieven, DBC-codering, taakbeheer, ordercreatie en invulling van zorginformatiebouwenstenen (ZIB's). De AI-oplossing fungeert daarbij als een digitale assistent die verslaglegging, administratieve verwerking en kwaliteitsbewaking overneemt. Hierdoor wordt tijd bespaard en verdwijnt de kans op gemiste registraties.

Verbreiding richting samenwerking

De toepassingen worden uitgebreid van de polikliniek naar SEH, IC-afdeling, direct aan het bed en op de OK. Daarnaast komt er ondersteuning voor telefonische en virtuele consulten (via bijvoorbeeld beeldbellen), headsetopnames en patiëntparticipatie. Ook worden nieuwe functies ontwikkeld, zoals een interne chatomgeving met GPT-functionaliteit. Zo groeit de AI van verslaglegger naar een integraal zorgcommunicatieplatform.

AI Agent

In de toekomst zien leveranciers zich als volwaardige *AI-agent* die zelfstandig administratieve, logistieke en klinische taken uitvoert, zodat zorgverleners zich volledig op patiëntenzorg kunnen richten.

Conclusie

Waar bij de vorige uitvraag met name de verbreiding van de technologie naar verslaglegging van andersoortige gesprekken in het ziekenhuis werd gezien, zien we nu een bredere *agentic* beweging, waarbij het EPD of andere systemen via spraak aangestuurd worden. Het zal nog een tijd duren voordat deze systemen geïntegreerd én conform de regels van o.a. de AI Act ingezet kunnen worden, maar de meeste leveranciers hebben een stip op de horizon met véél impact in de zorg.

4.8 Supportorganisatie

Aangezien de zorg een 24/7 sector is, is een goed functionerende supportorganisatie essentieel, vooral bij opschaling in ziekenhuizen. De volwassenheid van de supportstructuren varieert per leverancier:

- Volledige 24/7 ondersteuning: enkele partijen hebben al langdurige ervaring in ziekenhuizen en hanteren een volledig uitgerolde supportorganisatie met 24/7 bereikbaarheid. Ze werken met een Service Level Agreement (SLA), incidentprioritering en uitgebreide ondersteuning.
- Andere leveranciers hebben een helpdesk met ruime openingstijden en plannen uitbreiding naarmate hun technologie breder wordt ingezet, vooral in kritieke zorgomgevingen zoals de Spoedeisende Hulp.
- Sommige systemen hanteren automatische alerts die storingen detecteren en doorgeven, vooral bij samenwerkingen met huisartsenposten of andere 24/7-diensten.

- ITIL-gebaseerde service: een aantal aanbieders baseert hun support op best practices uit de IT-sector, zoals ITIL, om een gestructureerde en efficiënte dienstverlening te garanderen.

Conclusie

Over het algemeen hebben leveranciers met langere ziekenhuiservaring een robuustere supportorganisatie, terwijl nieuwere partijen hun ondersteuning nog verder ontwikkelen naarmate hun adoptie in ziekenhuizen groeit.

4.9 Prijsmodel

We zien dat prijsmodellen van leveranciers in ontwikkeling zijn: er is in deze markt nog veel mogelijk. Waar we in het vorige whitepaper de prijsmodellen in drie categorieën konden opsplitsen, lijkt er nu meer flexibiliteit in prijsmodellen. Dit past goed bij een markt in ontwikkeling.

We zien een tegenbeweging van 'tijdsbesparing' als belangrijkste manier om de licentiekosten op te vangen. Waar bij de vorige uitvraag de formule om de licentiekosten terug te verdienen vooral werd gekoppeld aan het zien van meer patiënten (of het wegwerken van wachtlijsten), zijn er nu drie richtingen:

- 1) Net zoals een halfjaar geleden leidt de inzet van de technologie tot een tijdsbesparing waardoor meer patiënten kunnen worden gezien (en daarmee wachtlijsten worden weggewerkt).
- 2) De technologie leidt tot een verbetering van registratie (van zorgactiviteiten) en daarmee een verhoging van de omzet.
- 3) De technologie zorgt voor een verlaging van werkdruk en een verhoging van werkplezier, en dat leidt tot een lager ziekteverzuim.

Het thema 'prijsmodel' is een belangrijk aandachtspunt, voor zowel de zorginstellingen als de techbedrijven. Want: leidt de inzet van deze technologie tot een efficiënter werkproces en wordt het ingezet om wachttijden te reduceren? Of wordt de technologie ingezet om werkplezier te verhogen, patiëntwaarde toe te voegen en werkdruk te verlagen? En wat is de balans tussen deze doelstellingen?

Conclusie

Leveranciers blijven in veel gevallen flexibel en ontwikkelen hun prijsstrategie in samenspraak met klanten. Er is nog geen prijsmodel dat de boventoon voert in Nederland.

5. Meetresultaten

In dit hoofdstuk delen we de meetresultaten van de live AI Validatielab-sessie van 6 oktober 2025. Hierbij gaat het om de snelheid van de modellen en om het opgeleverde resultaat. Longartsen, verpleegkundigen en verpleegkundig specialisten uit de mProve ziekenhuizen beoordeelden de gespreksverslagen.

5.1 Snelheid

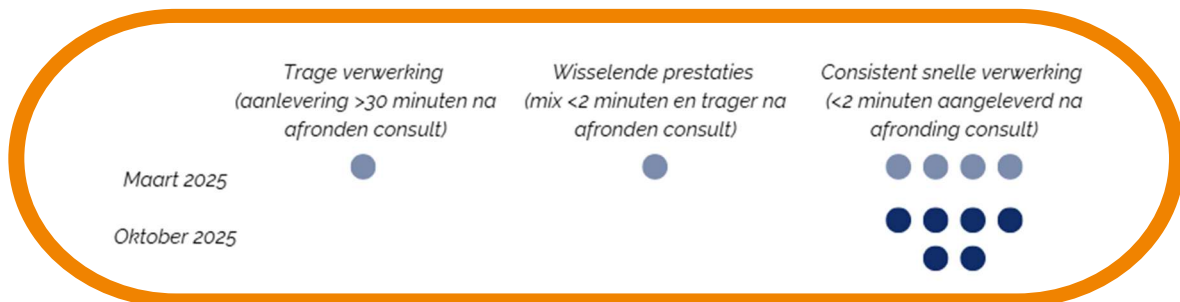
Voor een efficiënte workflow in de zorg is het cruciaal dat samenvattingen direct na een poliklinische afspraak beschikbaar zijn. De prestaties van de verschillende aanbieders op dit gebied variëren (figuur 4).

- **Snelle verwerking** (<2 minuten): de leveranciers hebben hun samenvatting binnen twee minuten gedeeld via het beschikbare mailadres. Het merendeel heeft de samenvattingen doorgaans binnen enkele seconden beschikbaar.
- **Wisselende prestaties**: deze categorie kwam niet meer voor tijdens deze sessie.
- **Trage verwerking** (>1 uur): deze categorie kwam niet meer voor tijdens deze sessie.

Dit onderdeel blijkt minder relevant dan de vorige keer: de performance van de systemen is significant beter dan in de februari-sessie. Alle leveranciers leverden direct na het afronden van het consult het verslag aan, zonder onverwachte vertraging.

Conclusie

Inmiddels presteren alle systemen goed genoeg om direct ondersteunend te zijn aan de workflow van de medisch specialist en andere zorgverleners. En dat is een randvoorwaarde om deze technologie goed geadopteerd te krijgen in de zorg.



Figuur 3: Snelheid verwerking

5.2 Toetsing casus longgeneeskunde

Aanpak

We hebben de poliklinische gesprekken opgenomen om achteraf een goede beoordeling te kunnen laten doen door medisch specialisten en een verpleegkundig specialist longgeneeskunde uit verschillende mProve ziekenhuizen.

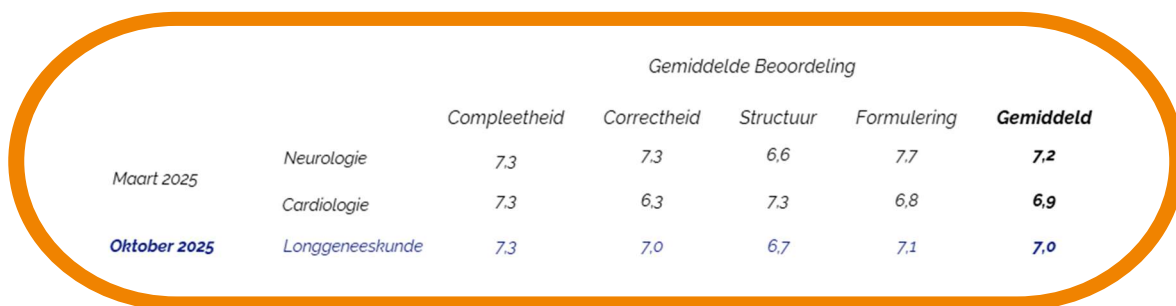
Zij ontvingen de opname, een document met de verslagen (zonder herleidbaarheid naar leveranciers in vorm of volgorde) én een beoordelingsformulier met ruimte om een score te geven op compleetheid, correctheid, structuur en formulering/zinsopbouw. Daarnaast

was er ruimte voor opmerkingen, een inschatting hoeveel tijd het zou kosten om het verslag te kunnen gebruiken in het EPD én welke tijdsbesparing er met het voorgelegde verslag gemoeid zou zijn (indien geïntegreerd in HiX).

We delen hierbij ook meteen onze limitatie bij deze praktijktest: het is één momentopname van een relatief kort gesprek (~5 minuten) tussen een medisch specialist, een patiënt en een mantelzorger. Het kan zijn dat de uitkomsten anders zijn bij langere consulten, bij een verbreding van het aantal beoordelaars, bij een ander specialisme of bij een consult zonder mantelzorger. Maar ook met deze kanttekening: het scala aan patiëntgesprekken in de spreekkamer is breed en we zoeken daarbij ook naar technologie die breed inzetbaar is.

Resultaten

Over de breedte scoren de medische verslagen een ruime voldoende. We zien daarin wel een bandbreedte: we zien zowel verslagen met een onvoldoende tot verslagen die zonder aanpassingen overgenomen kunnen worden in het EPD. Nu we deze sessie voor een tweede keer hebben uitgevoerd hebben we ook referentie-resultaten uit de eerste meting.



		Gemiddelde Beoordeling				
		Compleetheid	Correctheid	Structuur	Formulering	Gemiddeld
Maart 2025	Neurologie	7,3	7,3	6,6	7,7	7,2
	Cardiologie	7,3	6,3	7,3	6,8	6,9
Oktober 2025	Longgeneeskunde	7,3	7,0	6,7	7,1	7,0

Figuur 4: Evaluatiecriteria

De compleetheid is consistent over de sessies heen. Ook de gemiddelde score voor correctheid is een ruime voldoende. Rondom structuur zien we gemiddeld een voldoende. In theorie zou de structuurvraag goed in de technologie zelf opgelost kunnen worden als er voldoende draagvlak is voor één standaardrapportage bij medisch specialisten.

Als we inzoomen op individuele verslagen zien we wel een ruimere range van positieve en negatieve uitschieters. Net zoals de vorige keer zien we tevredenheidscijfers van 9,0 bij sommige samenvattingen, maar ook ruime onvoldoendes op andere gebieden bij andere verslagen. Dat beeld sluit aan bij de vorige keer dat we deze sessie hielden, de resultaten zijn soms nog grillig. We gaan graag in gesprek met de betreffende leveranciers om de grondoorzaak daarvan te achterhalen.

		Hoogste/Laagste per onderwerp			
		Compleetheid	Correctheid	Structuur	Formulering
Maart 2025	Neurologie	4,0-9,0	4,0-9,0	4,0-9,0	4,5-9,0
	Cardiologie	6,0-9,0	4,0-9,0	4,0-9,0	4,5-9,0
Oktober 2025	Longgeneeskunde	5,3-9,0	3,3-9,0	5,0-8,3	3,5-9,0

Figuur 5: Min/Maxscore

Compleetheid en relevantie

De beoordelaars vonden de verslagen overwegend compleet en inhoudelijk adequaat, maar merkten verbeterpunten op in leesbaarheid, structuur en duidelijkheid. Bij sommige verslagen werd in de voorliggende casus het hoesten van de patiënt gemist. Daarnaast miste soms de uitleg over het doel van de medicatie (afremmen van achteruitgang, niet klachtenverbetering), de vermelding van verkeerde patiëntverwachtingen en soms miste er een expliciete conclusie. Ook kwamen psychologische aspecten, bijwerkingen, zuurstofgebruik en emotionele lading in sommige gevallen niet altijd voldoende aan bod.

Correctheid en medische accuraatheid

De beoordelende specialisten beoordeelden de verslagen over het algemeen als correct, maar dat gold niet voor alle verslagen. Belangrijke informatie ontbrak soms, zoals interesse in lotgenotencontact of psychologische ondersteuning. Ook werden conclusies niet altijd onderbouwd of onjuist geformuleerd, bijvoorbeeld zonder verwijzing naar aanvullend onderzoek. Een voorbeeld voor het belang van medische accuratesse: bij één verslag werd het medicament Nintedanib geformuleerd als 'Nip-da-Nip', waarmee het vertrouwen in het verslag vanuit de beoordelaars in één keer gedecimeerd werd.

Structuur en herkenbaarheid in het EPD

Verslagen die aansluiten bij de gangbare structuur in HiX en medische verslaglegging werden beter beoordeeld. Sommige systemen waken hiervan af, bijvoorbeeld met onlogische kopjes of herhaling van patiëntinformatie onder verschillende secties. We vinden dat het verwerken van een verslag in een format dat aansluit bij de reguliere manier van rapporteren in een EPD een basis is voor leveranciers. Daarbij plaatsen we zelf de kanttekening dat die manier van structureren wel uniform omarmd moet worden door ziekenhuizen.

Formulering en leesbaarheid

De formulering in de verslagen werd wisselend beoordeeld. Sommige beoordelaars vonden de zinnen netjes opgebouwd en geschikt voor gebruik in bijvoorbeeld een brief, maar vaker werd gewezen op een staccato stijl, telegramachtige formuleringen en een gebrek aan hoofdletters, leestekens en vloeiende zinsopbouw, wat de leesbaarheid beperkte. Daarnaast werd het gebruik van medische termen zoals *fatigue* en *dyspneu* als afstandelijk ervaren; de voorkeur (van de betreffende respondent) ging uit naar patiëntgerichte taal. Enkele verslagen misten de beleving of emotie van de patiënt waardoor de toon soms kil overkwam. Samengevat werd de formulering vaak als

technisch correct, maar weinig natuurlijk en onvoldoende patiëntvriendelijk beoordeeld. Waarbij er natuurlijk ook stijlvoorkeur doorklonk in de beoordelingen.

Bijzonderheden

- Sommige aanbieders werken al aan personalisatie van samenvattingen, bijvoorbeeld met aanpasbare templates of instelbare samenvattingsniveaus. Dat is een mooie stap die ervoor zorgt dat het verslag aansluit bij de wensen van een bredere groep zorgprofessionals.
- Automatische registratie van zorgactiviteiten in het EPD wordt door sommige aanbieders al toegepast, wat administratieve lasten kan verminderen. Dit moet wel correct en controleerbaar zijn (en blijven).
- Een deel van deze beoordelingen gaat over de gewenste stijl van verslaglegging. Deze is niet uniform voor de specialisten van alle verschillende specialismen in een ziekenhuis. Er ligt daarom ook een taak voor vakgroepen of medische beroepsgroepen zelf om tot één gewenste standaard te komen.

Tijdsbesparing

Het was een kort consult dus veel tijdsbesparing werd niet mogelijk geacht. We zien in het mProve project 'Transcriptie in de Spreekkamer' vooral impact bij nieuwe patiënten met een lang consult. De casus betrof een herhaalpatiënt bij een kort consult.

Afhankelijk van de kwaliteit van het verslag zien we dat de minder beoordeelde verslagen 3 tot 5 minuten hersteltijd vergen en geen tijdsbesparing opleveren t.o.v. 'handmatig' rapporteren. De betere verslagen besparen medisch specialisten over het algemeen 3-5 minuten registratietijd bij correctiewerk van ~1 minuut. We zien dat laatste vooral terug bij leveranciers die op alle domeinen een beoordeling van een 7,5 of hoger scoorden.

Conclusie

De lat voor de bruikbaarheid van deze technologie ligt hoog, namelijk: 'kan dit verslag zonder aanpassingen (en met goedkeuring van de specialist) terechtkomen in het EPD'? We zien dat nog niet alle verslagen aan die standaard voldoen, maar een aanzienlijk deel ook wel.

De aanpassingen in structuur en formulering zijn voor een deel persoonlijke voorkeur van een zorgverlener, maar in essentie technisch relatief eenvoudig op te lossen. Daar ligt de verantwoordelijkheid voor doorontwikkeling bij zorgverleners, vakverenigingen en de ziekenhuizen zelf. Daarbij zouden vakverenigingen een rol kunnen pakken om te komen tot standaardisering van rapportagestructuur en -toon. Dat lijkt voor nu de meest impactvolle manier om een kwaliteitsverbetering op dit gebied te bewerkstelligen.

We zien dat de correctheid en medische accuraatheid bij sommige verslagen nog te wensen overlaten. Daarvoor zien we maar één oplossing: een continue feedbackloop met ziekenhuizen. Zo kan op structurele basis en grote schaal feedback worden gegeven en verwerkt in de tool. En wordt een nieuwe schaalsprong in kwaliteit gemaakt.

5.3 Toetsing verpleegkundige casus

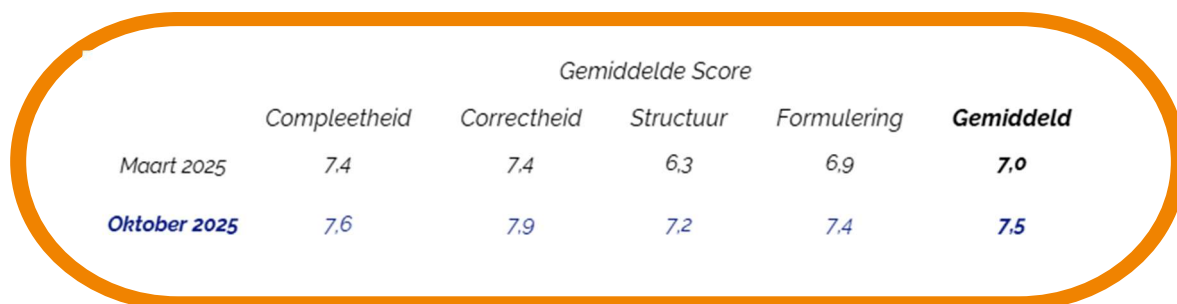
In tegenstelling tot onze voorgaande sessie werd dit keer vooraf aangekondigd dat er een verpleegkundige casus voorgelegd zou worden. Daardoor was er een betere voorbereiding op de casus mogelijk.

We hebben dezelfde methodologie gehanteerd als bij de medische casus. Het belangrijkste verschil: we hebben een handmatig geschreven verslag toegevoegd aan de te beoordelen verslagen. Op die manier proberen we de kwaliteit van de verslagen te relateren aan de dagelijkse praktijk waar de verslagen ook een bepaalde vorm en inhoud hebben. De beoordelaars zijn vooraf niet op de hoogte gesteld van het feit dat er ook een handgeschreven verslag toegevoegd zou worden.

In totaal hebben 6 verpleegkundigen de verslagen beoordeeld op compleetheid, correctheid, structuur en formulering/zinsopbouw. Tot slot is ook de vraag gesteld welke correctietijd nodig is om het verslag te kunnen gebruiken, en welke tijdsbesparing ermee gemoeid is.

Algemene beoordeling

In het algemeen zien we dat de gemiddelde beoordeling van de verpleegkundige verslaglegging een goede score oplevert. Ten opzichte van de meting in februari zien we op elk front een verbetering. Bij de gepresenteerde cijfers hieronder is het handmatig geschreven verslag geëxcludeerd, alleen 'artificiële' verslagen zijn toegevoegd. Hoewel we maar één verpleegkundige casus aan de leveranciers hebben voorgelegd zien we een duidelijke progressie.



	Gemiddelde Score				
	Compleetheid	Correctheid	Structuur	Formulering	Gemiddeld
Maart 2025	7,4	7,4	6,3	6,9	7,0
Oktober 2025	7,6	7,9	7,2	7,4	7,5

Figuur 6: Evaluatiecriteria

We zien dat er een aanzienlijke range van positieve en negatieve uitschieters is. Net zoals de vorige keer zien we zeer hoge tevredenheidscijfers op bepaalde fronten, hoewel we de 100% tevredenheidsscore niet aantikken. Tegelijkertijd zien we ook dat er dit keer minder verslagen door het ijs zakken. We zien daarmee een kwaliteitsverbetering in de verpleegkundige verslaglegging waar we tevreden mee zijn.

	Hoogste/Laagste per onderwerp			
	Compleetheid	Correctheid	Structuur	Formulering
Maart 2025	6,0-10,0	4,0-10,0	3,0-10,0	4,0-8,0
Oktober 2025	5,3-8,8	6,5-8,5	5,8-8,5	7,0-8,8

Figuur 7: Variatie in scores

Compleetheid en relevantie

De verpleegkundigen geven onvoldoendes aan verslagen die onvolledig of onevenwichtig zijn. In de voorliggende casus missen afspraken of onderdelen soms in de verslaglegging. Bijvoorbeeld betrokkenheid van maatschappelijk werk, hartfalenverpleegkundige, doseerbox of ontslagafspraken. Ook missen soms context en nuance uit het gesprek waardoor de essentie niet goed overkomt. Daarnaast zien we opmerkingen over irrelevante of privacygevoelige details (zoals het ziekteverleden van de dochter of het vermelden van leeftijd en basiskennis), wat de essentie vertroebelt. Sommige verslagen worden te summier of onduidelijk genoemd waardoor het lastig is om vervolgzorg te baseren op de tekst.

Bij hogere scores (8–10) waarderen verpleegkundigen vooral de volledigheid, duidelijke structuur en evenwichtige inhoud. Deze verslagen bevatten alle relevante elementen uit de casus, inclusief afspraken, beleid, diagnose en soms zelfs verpleegkundige reflectie of reacties van patiënt en familie. Ze worden omschreven als “mooi”, “uitgebreid” en “concreet”, met een goede combinatie van verhalende tekst en bulletpoints die de leesbaarheid in het dossier vergroot.

Hoewel sommige verslagen als lang worden ervaren, overheerst de indruk dat ze compleet en zorgvuldig zijn, met correcte follow-up en goed weergegeven beleid. De hoogste scores gaan naar verslagen die zowel volledig als overzichtelijk zijn, en waarin afspraken duidelijk, concreet en patiëntgericht zijn vastgelegd.

Correctheid en verpleegkundige accuraatheid

Bij minder beoordeelde verslagen geven verpleegkundigen aan dat deze onvolkomenheden bevatten in feitelijke juistheid of interpretatie. Er worden fouten genoemd in wie wat zegt of doet en sommige details worden verkeerd of onduidelijk weergegeven, zoals het gebruik van “baxterrol” in plaats van “medicatiebox” of doseerbox. Ook wordt opgemerkt dat sommige formuleringen nuance verliezen, bijvoorbeeld weerstand van familieleden die niet is terug te lezen. Verder komen kleine onnauwkeurigheden voor, zoals het vermelden van leeftijd of geslacht die niet benoemd werden.

Bij hogere scores geven verpleegkundigen aan dat de verslagen feitelijk kloppen, aansluiten bij de casus en een realistische weergave van de situatie en gemaakte afspraken bieden. Ze noemen de teksten “correct”, “heel nauwkeurig” en “goed aansluitend op het gesprek”. Medische, psychosociale en beleidsinformatie zijn in die verslagen goed verwerkt en de structuur zorgt voor helder inzicht in het zorgproces. De best

beoordeelde verslagen onderscheiden zich door hun consistentie, accuraatheid en natuurlijke aansluiting bij de praktijk. Ze geven de casus betrouwbaar en professioneel weer, zonder overbodige of foutieve details.

Structuur en herkenbaarheid

Bij minder beoordeelde verslagen geven verpleegkundigen aan dat de structuur van de verslagen vaak onduidelijk of onhandig is opgebouwd. Ze noemen dat sommige verslagen te lang zijn ("*twee A4'tjes voor één zorggesprek is teveel*"), te veel kopjes bevatten of juist vooral bestaan uit opsommingen zonder duidelijke samenhang. Dat maakt de tekst moeilijk te volgen en minder prettig leesbaar. In enkele gevallen wordt ook vermeld dat de opbouw onlogisch is of dat relevante informatie, zoals thuiszorgwensen en -plannen, verspreid staat in plaats van overzichtelijk bij elkaar. Verslagen met lage scores missen in de structuur een natuurlijke 'flow' en zijn niet direct bruikbaar voor opname in het EPD zonder significante aanpassing.

Bij hogere scores waarderen verpleegkundigen juist de duidelijke, logische en overzichtelijke opbouw. Een heldere indeling in onderdelen zoals anamnese, beleid, diagnose en follow-up wordt als prettig ervaren, net als het gebruik van bulletpoints en gescheiden secties voor beleid of actiepunten. Verslagen met een sterke structuur worden omschreven als goed leesbaar, professioneel en makkelijk te verwerken in het EPD. Ook het combineren van een verhalend deel met een kort actieoverzicht wordt als effectief gezien.

Formulering en leesbaarheid

Bij de verslagen die minder scoorden, waren verslagen te langdradig of juist te beknopt waardoor de balans tussen volledigheid en leesbaarheid ontbrak. Ze noemen dat sommige teksten "een heel boekwerk" vormen met lange zinnen en herhalingen, terwijl andere verslagen juist te kort door de bocht zijn en essentiële informatie missen, zoals afspraken over thuiszorg. Hoewel de verpleegkundige taal meestal correct is, ervaren sommigen de formulering als stug of onnatuurlijk lopend.

Bij hogere scores waarderen verpleegkundigen vooral de duidelijkheid, professionele toon en goede zinsopbouw. Verslagen die kort en krachtig zijn, worden als prettig en goed leesbaar ervaren. De taal wordt dan omschreven als zakelijk, helder en professioneel, met een natuurlijke opbouw die logisch en overzichtelijk leest.

Tijdsbesparing

Nieuw na deze sessie is het beoordelen van (potentiële) tijdsbesparing doordat het verslag automatisch gegenereerd wordt. Bij integratie van deze technologie in HiX wordt de volgende tijdsbesparing gerapporteerd:

- Om de aangeleverde verslagen te kunnen gebruiken in het EPD zou gemiddeld nog 3,4 minuten correctiewerk nodig zijn.
- Tegelijkertijd wordt bij 22% van de verslagen een tijdsbesparing van >10 minuten verwacht, in 44% van de gevallen 5-10 minuten, in 17% van de gevallen een tijdsbesparing van 3-5 minuten en 17% van de gevallen 0-2 minuten tijdsbesparing.
- De gemiddeld verwachte tijdsbesparing van de verslagen na het doorvoeren van aanpassingen is 6,3 minuten.

Onder de streep tonen we hiermee aan dat er een tijdsbesparing optreedt bij de inzet van deze technologie door verpleegkundigen bij EVV-verslagen. Daarbij is de netto

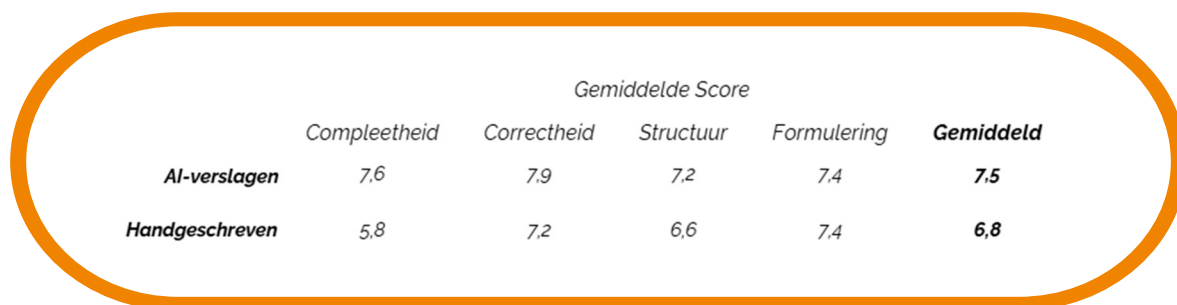
tijdsbesparing in sommige gevallen bijna 8 minuten, inclusief het aanpassen van het automatisch gegenereerde gespreksverslag.

Discussie referentiekader

Het toevoegen van een handmatig geschreven verslag levert een belangrijk discussiepunt op voor nieuwe technologie in de (medisch specialistische) zorg. Want: beoordelen we technologie op een wijze waarbij we naar nul incidenten en een perfecte score streven? In de huidige werkwijze komen ook incidenten voor, mensen zijn niet feilloos.

Verpleegkundigen van mProve ziekenhuis Noordwest Ziekenhuisgroep leverden een belangrijke bouwsteen aan voor dit onderdeel van ons onderzoek. In een setting met vijf verpleegkundigen hebben zij hetzelfde EVV-gesprek gevat in een verslag voor in het EPD. Eén van die verslagen is aangeleverd voor beoordeling door de verpleegkundigen die meededen aan deze sessie van het 'AI Validatielab'.

En dat levert boeiende resultaten op. We zien dat het handmatig geschreven verslag kort en bondig is, en erg herkenbaar wordt beoordeeld qua formulering. De AI-verslagen scoren daar gelijkwaardig aan. We zien echter dat het handmatige verslag dermate kort en bondig is opgesteld dat de beoordelend verpleegkundigen het té kort vinden en daardoor mindere scores geven aan compleetheid.



	Gemiddelde Score				
	Compleetheid	Correctheid	Structuur	Formulering	Gemiddeld
AI-verslagen	7,6	7,9	7,2	7,4	7,5
Handgeschreven	5,8	7,2	6,6	7,4	6,8

Figuur 8: Gemiddelde scores verpleegkundige casus

Bijzonderheden

Bij deze sessie waren leveranciers wel voorbereid op een verpleegkundige casus. En dat zien we direct terug: de structuur van de gesprekken werd beter beoordeeld dan bij vorige sessie.

Conclusie

We zijn zeer tevreden over de verbeteringen in de verpleegkundige verslagen ten opzichte van de sessie in maart 2025. We zien op alle fronten een vermindering van *outliers* als het gaat om overdreven goede of slechte verslagen. We zien dat het gemiddelde verslag en zeker ook de betere verslagen, met minimale aanpassingstijd bruikbaar kunnen zijn in de klinische praktijk.

Verpleegkundigen beoordeelden dat 5 van de 6 leveranciers betere verslagen leverden dan het handmatig geschreven verslag. We kunnen niet genoeg benadrukken hoe

belangrijk deze bevinding is: binnen mProve alleen al zien we dat 10.500 verpleegkundigen baat kunnen hebben bij deze technologie.

En dat is belangrijk! In arbeidsmarktdoorrekeningen zien we dat het toekomstige tekort aan verpleegkundigen in de toekomst in absolute aantallen circa 6x groter is dan het tekort aan medisch specialisten. Dat de technologie in de huidige vorm al kwalitatief goede verslagen kan opleveren die (registratie-)tijd besparen geeft ons vertrouwen voor de toekomst.

6. Takeaways

Generatieve AI zet steeds meer stappen om uit te groeien tot '*ambient*' technologie die ongemerkt op de achtergrond 'meeluistert' en verslaat wat er gebeurt. Deze tweede live 'AI Validatielab'-sessie door het mProve Funxlab heeft opnieuw waardevolle inzichten opgeleverd over de huidige stand van AI-gestuurde samenvattingen in de zorg en de ontwikkelrichting van leveranciers.

De richting is veelbelovend en we zijn optimistisch over de stappen die worden gezet. De technologie ontwikkelt snel, maar heeft op dit moment soms nog onvoldoende grip op nuances en context. En dat leidt tot correctiewerk voor betrouwbare verslaglegging. Het grootste deel van de verslagen leidt nu al tot een verlaging van de administratieve druk. We kunnen deze conclusie op dit moment echter nog niet generaliseren naar alle verslagen.

We zien een verbreding van de inzet van deze technologie: we zien dat de verslagen die geschreven zijn voor verpleegkundige gesprekken met patiënten inmiddels erg goed beoordeeld worden en leiden tot een potentiële tijdsbesparing op administratietijd voor een steeds groter wordende groep zorgverleners. We bevelen ziekenhuizen dan ook aan om pilots en opschaling richting verpleegkundigen en andere zorgverleners te verkennen. De potentiële impact is ook buiten de spreekkamer groot. En waar de focus op de medisch (of verpleegkundig) specialist ligt: de grootste potentie zit bij de specialismen met langere consulten.

We zien daarnaast een verbreding van de technologie richting Agentic AI. We zijn absoluut fan van het idee om AI als collega te kunnen inzetten in de zorg, en van de bijbehorende beloften om de administratietijd terug te dringen. Maar we roepen de betrokken tech-bedrijven ook op om naar de Medical Device Regulation (MDR) te kijken. Roadmaps van leveranciers zitten in het grijze gebied van de MDR. Dergelijke trajecten zijn tijdrovend en van de lange adem: wees daarop voorbereid.

Verder zien we in ons onderzoek dat de integratie van deze technologie in het elektronisch patiëntendossier weinig nieuwe mijlpalen heeft opgeleverd afgelopen halfjaar. Tegelijkertijd is deze integratie ongelofelijk belangrijk voor de reductie van administratietijd en opschaling van de technologie. De Nederlandse zorg kan nú verbeteringen gebruiken die transformatief van aard zijn, dit stukt in Nederland nog.

We zien ook dat zorginstellingen een rol hebben in het verbeteren van de kwaliteit van deze technologie. Naast het geven van terugkerende feedback over de kwaliteit van de producten, zodat deze kunnen verbeteren, is standaardisatie van rapportage ook een vraagstuk. Welke structuur heeft een poliklinisch verslag idealiter? Welk taalgebruik hoort daarbij? Kortom: wat is de standaard?

Conform het Europese beeld zien we een grote afhankelijkheid van Amerikaanse techreuzen voor zowel taalmodellen als datacenters. Hier zien we op dit moment ook niet veel keuzemogelijkheid in. Globalisering levert een afhankelijkheid op die ook te ver kan gaan: op welke manier pakken we de regie op dit onderwerp? Het begint in ieder geval bij bewustzijn.

Ondertussen blijft het mProve project 'Transcriptie in de Spreekkamer' doorgaan: inmiddels maken zorgverleners in 6 van de 7 mProve ziekenhuizen gebruik van deze technologie. Het 7^e ziekenhuis is onderweg. Met onze samenwerking voorkomen we dat elk ziekenhuis afzonderlijk het wiel opnieuw uitvindt en zetten we onze innovatiecapaciteit zo slim en efficiënt mogelijk in. We delen onze bevindingen ook nadrukkelijk buiten mProve, zodat we ook dáár bouwstenen leveren voor de inzet van deze technologie. De noodzaak hiertoe wordt met de dag groter.

Stay tuned. **Het wordt Fun(x)** in de Spreekkamer (en vër daarbuiten)!

Over mProve

Vanuit de kracht van 7 ambitieuze topklinische ziekenhuizen: Albert Schweitzer ziekenhuis, Isala, Jeroen Bosch Ziekenhuis, Noordwest Ziekenhuisgroep, Máxima MC, Rijnstate en Zuyderland Medisch Centrum, geven wij de zorg van de toekomst vorm. We versterken elkaar door van elkaar te leren, resultaten te delen en de beste aanpak over te nemen. Als aanjager van vernieuwing ontwikkelen wij doeltreffende, schaalbare innovaties. Nederland heeft slimme oplossingen nodig en wij realiseren ze.

mProve. Samen voor de toekomst

mProve Funxlab

Eén van de grootste uitdagingen in de zorg is het tekort aan personeel terwijl de zorgvraag toeneemt. Met het mProve Funxlab focussen we op technologische innovaties die het werk in de ziekenhuizen leuker en makkelijker maken... Slimme technologie met een functie én met de funfactor.

De 7 mProve ziekenhuizen delen binnen het mProve Funxlab hun ervaringen en resultaten met innovaties die zorgverleners helpen om efficiënter, effectiever en met meer plezier te werken. Bijvoorbeeld technologie die administratieve taken overneemt. Wat in de praktijk goed werkt, wordt via Funxlab verder geïmplementeerd en opgeschaald in de deelnemende ziekenhuizen. Ook scouten we actief in de markt om technologie die op het punt van doorbreken staat, snel te kunnen adopteren en samen in te kopen.

We vernieuwen in het Funxlab de zorg met en vóór zorgverleners. Als zij hun werk goed kunnen doen, leidt dat ook tot betere zorguitkomsten. Met onze werkwijze zorgen we er bovendien voor dat niet alle ziekenhuizen pilots blijven doen met dezelfde technologische oplossingen. Want, je kan de verandercapaciteit die de zorg heeft maar één keer inzetten.

Colofon

Copyright © 2025 mProve

Auteur: Jos Knaapen

Met speciale dank aan: Janneke van Strien, Gijs Vanstreels, Suzanne Anbeek, Stefan Kroese, Hanneke van Heijst – van Doremalen, Bernadette van Kemenade, Lucyl Verhoeven, Marinka Stierp, Daniëlle Zwagerman, Miranda Bergsma, Rochelle Sigterman Nelissen, Mariëlle Lutke Willink -Smets, Ben Tomlow, Esther Verstraete, Tom van de Wielen, Jorik Slotman, Wouter Blox, Pathman Pasupathy, Thymo Griesenbrock, Tamara Jonker, Mark Meijer *en vele anderen*.

Datum: November 2025

Contact: Funxlab@mprove.nu

www.mprove.nu

mProve
funxlab!