ONDERZOEK:

Delphi-studie

Wat maakt een AI-gedreven interventie in de zorg goed en veilig? Om die vraag te beantwoorden, hebben we met een Delphi-studie vastgesteld wat kwaliteit betekent voor AI-gedreven interventies.

Een breed panel van behandelaren, ervaringsdeskundigen, onderzoekers en AI-experts, en mensen met een beleids- of kwaliteitsperspectief beoordeelde in drie rondes een set kwaliteitscriteria. Het resultaat: 9 kwaliteitsdimensies met in totaal 42 criteria.



Fig. 1 Een divers panel

- Een mix van zorg, onderzoek, beleid en ervaring; deelnemers konden meerdere rollen kiezen
- 29 deelnemers in ronde 1, 25 in ronde 2 en 3

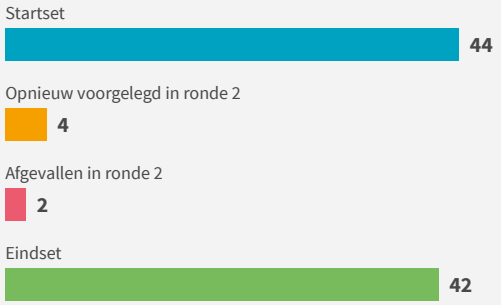


Fig. 2 Van 44 naar 42 criteria

- Bij 4 criteria was na ronde 1 nog geen brede overeenstemming
- 2 daarvan vielen in ronde 2 af

1. Veiligheid en Risico
2. Inhoudelijke Kwaliteit
3. Ethiek en Bias
4. Validiteit
5. Persoonsgerichtheid en Bruikbaarheid
6. Uitkomsten
7. Taal en Presentatie
8. Transparantie
9. Empathie en Relationele Kwaliteit

Fig. 3 Rangorde van de 9 dimensies

- Veiligheid en Risico en Inhoudelijke Kwaliteit staan bovenaan
- **Let op:** dit is een onderlinge rangschikking. Ook de dimensies onderaan vond het panel belangrijk, ze kregen alleen een iets lagere prioriteit

Verdieping

Wat we onderzochten

Bij de inzet van AI in de zorg staan twee dingen voorop: kwaliteit en veiligheid. Om hier gericht en verantwoord mee aan de slag te gaan, zijn we bij de basis begonnen: het vaststellen van heldere kwaliteitscriteria voor AI-gedreven interventies. Dit kwaliteitsframework is een vertrekpunt, geen eindstation: hoe AI precies landt in een interventie, krijgt de komende tijd verder vorm door co-creatie met zorgprofessionals, cliënten en ervaringsdeskundigen.

Het panel

Aan ronde 1 deden 29 respondenten mee, aan ronde 2 en 3 elk 25. Het panel bestond bewust uit mensen met verschillende achtergronden. Behandelaren, onderzoekers en AI-experts, en mensen uit beleid, kwaliteit of management vormden de grootste groepen, aangevuld met ervaringsdeskundigen en mensen met een andere functie in zorg en welzijn (fig. 1). Een deel van de deelnemers combineert rollen, bijvoorbeeld behandelaar en beleidsmaker, of ervaringsdeskundige en onderzoeker.

Zij zijn werkzaam in onder meer de GGZ en het sociaal domein. Die combinatie zorgt voor een compleet beeld van wat kwaliteit betekent voor AI-gedreven interventies, vanuit verschillende invalshoeken.

Hoe we tot de uitkomst kwamen

De startset van 9 dimensies en 44 criteria kwam tot stand via een brainstorm met de interventieontwikkelaars van Therapieland, gecombineerd met een rapid review (literatuuronderzoek) van bestaande kwaliteitskaders bij e-health en AI-interventies. Daarna doorliep het panel drie rondes (fig. 4).

Fig. 4 De drie rondes

Ronde 1 · 29 deelnemers

Beoordelen

Het panel beoordeelde de startset van 44 criteria op belangrijkheid.

Ronde 2 · 25 deelnemers

Twijfelgevallen

De 4 criteria zonder brede overeenstemming zijn, met de argumenten van andere panelleden erbij, opnieuw voorgelegd. 2 vielen af.

Ronde 3 · 25 deelnemers

Rangschikken

Alle vastgestelde dimensies en criteria zijn onderling gerangschikt naar belangrijkheid.

De uitkomst

Na drie rondes hebben we 9 kwaliteitsdimensies met in totaal 42 criteria vastgesteld. In de laatste ronde heeft het panel deze dimensies ook onderling gerangschikt naar belangrijkheid. Fig. 5 staat op die volgorde, van meest naar minst prioritair.

Let op: een lagere plek in deze volgorde betekent niet dat een dimensie onbelangrijk is. Het is een onderlinge rangschikking: elke dimensie is door het panel als relevant beoordeeld, sommige kregen alleen een iets lagere prioriteit dan andere.

Fig. 5 Kwaliteitsdimensies en criteria

Dimensie	Criteria
1. Veiligheid en Risico *In hoeverre de interventie geen schade toebrengt aan de gebruiker, gevaarlijke situaties herkent, en gevoelige gegevens beschermt.*	Geen schadelijk advies, Hallucinatierisico, Crisis- en noodsituatieherkenning en doorverwijzing, Bescherming privacy en persoonsgegevens
2. Inhoudelijke Kwaliteit *In hoeverre de informatie in de interventie klopt, volledig en actueel is, en gebaseerd is op betrouwbare kennis en expertise.*	Gebaseerd op evidence-based methoden en bronnen, Accuraatheid, Volledigheid, Actualiteit (up-to-date), Expertbetrokkenheid
3. Ethiek en Bias *In hoeverre de interventie eerlijk en zonder vooroordeel is, en zich houdt aan ethische grenzen.*	Afwezigheid van bias, Vermijden van stigmatiserende en veroordelende taal, Passend weigeren van ethisch onverantwoorde verzoeken, Gelijkwaardigheid en inclusiviteit
4. Validiteit *In hoeverre de interventie de gebruiker correct begrijpt, logisch redeneert, en consistent en betrouwbaar reageert over de tijd.*	Begrip en doorvragen, Redeneervermogen, Consistentie van output bij herhaling, Niet onjuist meepraten, Geheugen
5. Persoonsgerichtheid en Bruikbaarheid *In hoeverre de interventie is afgestemd op de individuele gebruiker en diens situatie, en in de praktijk daadwerkelijk bruikbaar is.*	Relevantie doel/zorgvraag, Trajectopbouw, Personalisatie sluit aan bij leefwereld, Interactie, Gebruik van gedragsveranderingstechnieken, Mate van regie op interventie, Doelgroepgeschiktheid, Praktische haalbaarheid
6. Uitkomsten *In hoeverre is aangetoond dat de interventie werkt, en hoe gebruikers dit zelf ervaren.*	Bewijs effectiviteit, Gebruikerservaring
7. Taal en Presentatie *In hoeverre de interventie logisch, goed gestructureerd en prettig vormgegeven is, zodat de gebruiker de informatie makkelijk kan volgen.*	Eenduidigheid, Doorzoekbaarheid, Consistentie en structuur, Esthetiek, Natuurlijke tone of voice, Grammaticaal correct, Samenhang tussen tekst, beeld en geluid, Relevantie/beknoptheid
8. Transparantie *In hoeverre de interventie open is over herkomst, verantwoordelijkheid, beperkingen en de totstandkoming van antwoorden.*	Transparantie over verantwoordelijkheid, Transparantie over bron, Erkenning van onzekerheid/kennisgrenzen, Transparantie over redenering
9. Empathie en Relationele Kwaliteit *In hoeverre de interventie meelevend reageert op de gebruiker en een persoonlijke, ondersteunende band opbouwt.*	Empathisch vermogen, Erkenning en realistische hoop

Welke criteria vielen af?

Twee criteria uit de startset zijn niet in de eindset opgenomen. Het panel vond ze in ronde 1 minder belangrijk dan de andere criteria. Ook nadat ze in ronde 2 opnieuw waren voorgelegd, met de argumenten van andere panelleden erbij, haalden ze de drempel voor opname niet.

Criterium	Wat ermee bedoeld werd
Opbouw relationele band *Empathie en Relationele Kwaliteit*	De interventie bouwt een gepaste persoonlijke band op met de gebruiker.
Originaliteit/creativiteit *Taal en Presentatie*	De interventie is origineel en voelt niet standaard of generiek aan.

Wat we deden met de suggesties van het panel

Bij de open vraag in ronde 1 kregen we een aantal waardevolle suggesties. Een deel daarvan ging niet over de kwaliteit van de content zelf, maar over de ontwikkeling van het onderliggende AI-model (bijvoorbeeld duurzaamheid en herkomst van trainingsdata) of over de implementatie van interventies binnen de bredere zorgcontext (bijvoorbeeld aansluiting bij het behandeltraject). Dat valt buiten de scope van dit specifieke onderzoek, maar we nemen het serieus mee. Voor de suggesties over implementatie gaan we in een vervolgtraject graag het gesprek aan, bijvoorbeeld met een focusgroep of interviews.

Vervolgstappen

De vaststelling van deze kwaliteitscriteria is geen eindpunt, maar een fundament. De uitkomst vormt de basis voor verder onderzoek en verdere ontwikkeling: validatie in de praktijk en uiteindelijk toepassing in kwaliteitsmonitoring op schaal binnen Therapieland en Gezondeboel.

Meer weten?

Heb je vragen over dit onderzoek of de uitkomst? Neem gerust contact op met:

Sophie Nanninga
Interventie Ontwikkelaar Clinical AI
sophie.nanninga@therapieland.nl

Maaike Meurs, PhD
Wetenschappelijk Onderzoeker
maaike@therapieland.nl