

AI-beslisondersteuning in de zorg

De waarde van Large Language Models voor jeugdartsen: een verkenning

Beslisondersteunende systemen met AI bieden kansen voor de jeugdgezondheidszorg, maar botsen met de praktijk waarin richtlijnen tekstueel complex zijn en dossiers grotendeels uit vrije tekst bestaan. In dit onderzoek wordt verkend in hoeverre AI-adviezen overeenkomen met de richtlijnen en het oordeel van jeugdartsen.

Imme Moeskops, Tim van den Broek, Arjan Huizing en Olivier Blanson Henkemans

Beslisondersteunende systemen op basis van kunstmatige intelligentie (AI) bieden nieuwe mogelijkheden om de kwaliteit, consistentie en doelmatigheid van zorg te verbeteren. In toenemende mate wordt generatieve AI, zoals Large Language Models (LLMs), onderzocht als middel om zorgprofessionals te ondersteunen bij het analyseren van informatie, het interpreteren van protocollen en richtlijnen, en het formuleren van beleid. Deze technologie biedt snellere toegang tot medische kennis en ondersteunt klinisch redeneren bij onzekerheid of tijdsdruk (Shortliffe & Sepúlveda, 2018).

De implementatie van AI in de klinische praktijk blijkt echter complex. Succesvolle inzet vraagt niet alleen om technische betrouwbaarheid, maar ook om afstemming met werkroutines, professionele autonomie en gevoeligheid voor context (Liu et al., 2023). In dit licht groeit de aandacht voor *explainable* en *trustworthy* AI, waarbij transparantie, herleidbaarheid en gebruiksvriendelijkheid centraal staan (Fjeld et al., 2020). Generatieve modellen voegen hieraan toe dat ze betekenis kunnen ontleenen aan ongestructureerde data, zoals vrije tekst uit dossiers of patiëntvragen, en zo nieuwe vormen van interactie met protocollen en richtlijnen mogelijk maken (Wang et al., 2023; Hoekstra et al., 2024).

In dit artikel verkennen we de inzet van AI als beslisondersteuning binnen de jeugdgezondheidszorg (JGZ). Aan de hand van casussen onderzoeken we in hoeverre een LLM beleidsadviezen kan genereren die zowel aansluiten bij geldende JGZ-richtlijnen als bij het professionele oordeel van jeugdartsen. Deze verkenning levert inzichten op die breder relevant zijn voor de inzet van AI in klinische besluitvorming, waar het samenspel tussen model, richtlijn en professional cruciaal is.

De aanleiding ligt in het gebruik van bestaande hulpmiddelen binnen de JGZ, zoals de Slimme Richtlijn Module (SRM). Deze module ondersteunt professionals bij het genereren van richtlijnconform beleid via beslisbomen en knooppunten, en draagt bij aan uniformiteit in dossierregistratie (Blanson Henkemans et al., 2021). Tegelijkertijd blijkt dat vrije tekstinput moeilijk te structureren is en buiten de reikwijdte van deze SRM valt. Ook zijn JGZ-richtlijnen vaak omvangrijk en tekstueel complex, waardoor ze lastig zijn te vangen in afgebakende logica. Dit maakt LLMs, die juist getraind zijn op het begrijpen en bevragen van vrije tekst, een relevante aanvulling voor beslisondersteuning (Lekadir et al., 2023). De centrale onderzoeksvraag luidt: in hoeverre genereert een AI-model beleidsadviezen die aansluiten bij zowel de geldende richtlijnen als het professionele oordeel van jeugdartsen, en welke factoren verklaren eventuele verschillen?

lijnen vaak omvangrijk en tekstueel complex, waardoor ze lastig zijn te vangen in afgebakende logica. Dit maakt LLMs, die juist getraind zijn op het begrijpen en bevragen van vrije tekst, een relevante aanvulling voor beslisondersteuning (Lekadir et al., 2023). De centrale onderzoeksvraag luidt: in hoeverre genereert een AI-model beleidsadviezen die aansluiten bij zowel de geldende richtlijnen als het professionele oordeel van jeugdartsen, en welke factoren verklaren eventuele verschillen?

Methode

Deze studie is opgezet als eerste verkenning van de toepassingsmogelijkheden van generatieve AI, inclusief LLM, binnen een domein waarin richtlijnen al een centrale rol spelen. De JGZ diende hierbij als toepassingscontext, met als focus kinderen van 0 tot 4 jaar, omdat in deze leeftijdsgroep veel standaardcontactmomenten plaatsvinden en richtlijngebruik sterk is ingebed in de dagelijkse praktijk. Dit maakt de JGZ een geschikte setting om de aansluiting van AI-gegenereerd advies op bestaande richtlijnsystematiek én klinisch redeneren te onderzoeken.

Het AI-model

In dit onderzoek werd een AI-systeem ingezet met als doel om informatie over kinderen uit het digitale dossier systematisch te interpreteren aan de hand van actuele richtlijnen. Het systeem doet dat door drie technologieën te combineren:

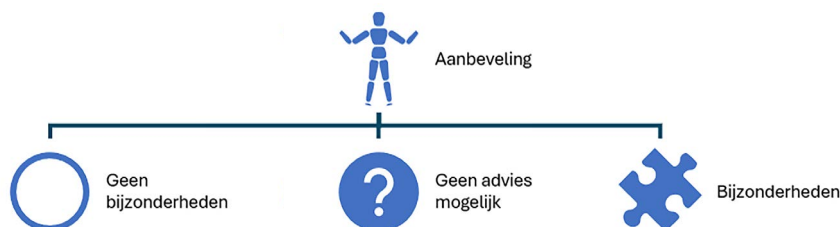
- gestructureerde gegevensverwerking: het systeem weet welke gegevens belangrijk zijn;
- slimme zoektechnologie (semantisch zoeken): het systeem begrijpt waar de informatie over gaat, niet alleen de exacte woorden;
- taalmodellen (LLM's): het systeem kan gewone taal begrijpen en samenvatten, zoals een mens dat ook zou doen.

De input voor het systeem bestaat uit informatie die is vastgelegd volgens de *Basisdataset JGZ* (BDS): een vaste indeling voor gegevens in de JGZ, zoals geboortedatum, lichamelijke observaties of gespreksverslagen. Op basis van vooraf vast-

gelegde koppelingen weet het systeem welke gegevens horen bij welke richtlijn. Stel bijvoorbeeld dat een arts bij een baby een hartruis hoort en dat noteert, dan weet het systeem: dit hoort bij de richtlijn *Hartafwijkingen*. Die koppeling is vooraf ingesteld. Vervolgens vat het AI-systeem de informatie uit het dossier samen in begrijpelijke taal. Deze samenvatting vormt de basis voor de volgende stap: het zoeken naar passende informatie in de richtlijn. Om dat goed te doen is de richtlijn van tevoren opgedeeld in kleine tekstfragmenten (ook wel *chunks* genoemd), en zijn die slim opgeslagen. Elk fragment is vertaald naar een soort 'numerieke vingerafdruk' (embedding), zodat het systeem snel kan bepalen welke stukken inhoudelijk passen bij wat er in het dossier staat. Daarbij bewaart het systeem ook metadata, zoals het paginanummer van elk fragment. Per bevinding in het dossier zoekt het systeem zo de meest passende richtlijnfragmenten. Die fragmenten worden één voor één beoordeeld door een kleiner taalmodel dat bepaalt of ze echt inhoudelijk relevant zijn. Alleen de stukken die goed aansluiten, worden uiteindelijk meegenomen in de laatste analyse. In die laatste stap genereert een krachtiger taalmodel een inhoudelijke interpretatie en conclusie. Daarin staat bijvoorbeeld of er een bijzonderheid is, welke vervolgstappen nodig zijn (bijvoorbeeld: doorverwijzen), en op welke pagina van de richtlijn dat advies gebaseerd is.

Casusverzameling en classificatie

Er werd gewerkt met vijftien kind-casussen. Negen hiervan zijn verzameld bij artsen in opleiding tot specialist (aios-sen), waarbij gebruik is gemaakt van een standaardformat voor het kopiëren van relevante onderdelen uit het digitale dossier. Vier aanvullende casussen zijn ontwikkeld door de onderzoeker, eveneens een aios, en twee zijn ingebracht door een jeugdarts. Elke casus was volgens hetzelfde format opgesteld (algemene beschrijving van het kind, weergave in het digitale dossier en een aanbeveling door de jeugdarts) en bevatte voldoende informatie om op een realistische manier een zorginhoudelijke beslissing te kunnen nemen – zoals het wel of niet inzetten van vervolgonderzoek, extra contactmomenten of een verwijzing.



Afbeelding 1. Classificaties van aanbevelingen: Geen bijzonderheden, Geen advies mogelijk (met beschikbare data), Bijzonderheden.

De casussen werden geclassificeerd binnen vier medische domeinen waarvoor heldere JGZ-richtlijnen bestaan: huidafwijkingen, heupdysplasie, voeding en eetgedrag, en hartafwijkingen. Per domein is een onderscheid gemaakt tussen (1) de aan- of afwezigheid van bijzonderheden (afbeelding 1) en (2) in het geval van bijzonderheden, de geadviseerde vervolgstappen (afbeelding 2).

Beleidsadviezen: drie bronnen

Voor elke casus zijn drie beleidsadviezen geformuleerd:

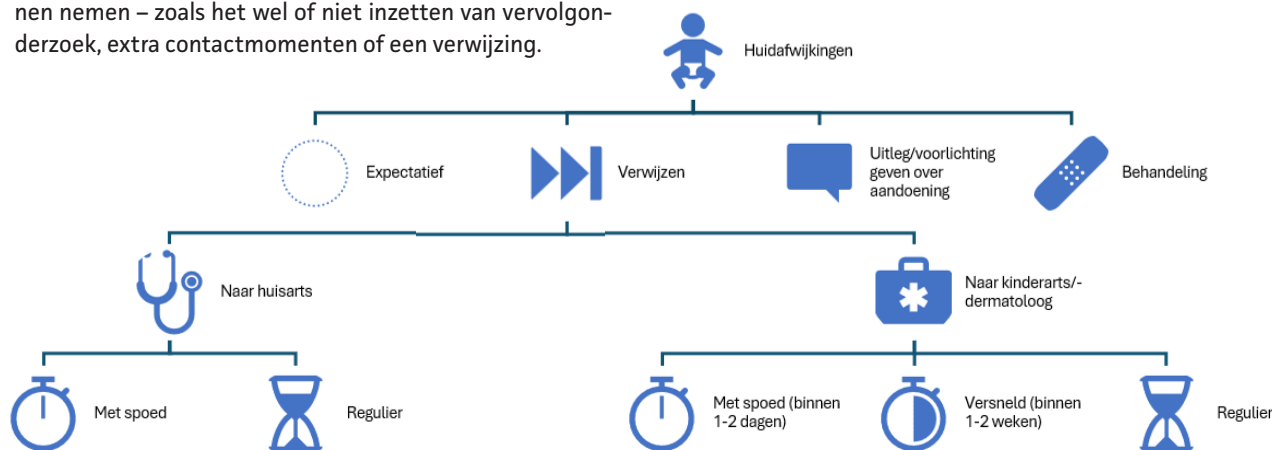
- professioneel advies: het originele beleid zoals vastgesteld door de jeugdarts;
- richtlijnadvies: het advies op basis van de geldende JGZ-richtlijn;
- AI-advies: output van de LLM.

Het AI-model werd vijfmaal per casus geraadpleegd met dezelfde prompt, om de consistentie van de output te evalueren. De prompts werden zorgvuldig opgebouwd met verwijzing naar het richtlijndomein, de relevante classificaties (zoals 'geen bijzonderheden'/'wel bijzonderheden'), en de instructie om een kort beleidsadvies te genereren. Waar nodig werd de prompt licht aangepast om de relevante context toe te voegen, zonder het model te sturen richting een bepaald antwoord.

Analyse

De drie adviezen per casus zijn vergeleken op:

1. Classificatie: komt de beoordeling 'bijzonderheden aanwezig of niet' overeen tussen model, richtlijn en arts?
2. Vervolgstappen: is het geadviseerde beleid in lijn qua vervolgstappen en interventies?



Afbeelding 2. Vervolgstappen in het geval van classificatie Bijzonderheid. Hier geïllustreerd voor de casus kind met een huidafwijking.

Bij afwijkingen werd gekeken naar mogelijke verklaringen, zoals ontbrekende informatie, regionale afspraken, of professionele inschatting gebaseerd op ervaring. Op deze manier werd niet alleen de overeenstemming in kaart gebracht, maar ook de aard en relevantie van eventuele verschillen geanalyseerd.

Resultaat

Overeenstemming met richtlijn

Tabel 1 toont per richtlijndomein de mate van overeenstemming tussen het AI-model, de richtlijn en het beleid van de jeugdarts. De analyse is opgesplitst in (1) classificatie (aan- of afwezigheid van bijzonderheden) en (2) de geadviseerde vervolgstappen. In totaal zijn 36 casussen geanalyseerd, verdeeld over vier domeinen. De classificatie van het AI-model kwam bij alle casussen overeen met de richtlijn en in 35 van de 36 casussen overeen met die van de arts (97 procent). Bij de vervolgstappen was de overeenstemming tussen het AI-model en de richtlijn in totaal 78 procent, en met de jeugdarts 72 procent. De meeste afwijkingen in vervolgstapadvies deden zich voor bij huidafwijkingen (in 5 van de 15 casussen) en heupdysplasie (in 4 van de 15 casussen), tussen de arts en het AI-model evenals tussen de arts en de richtlijn. Deze verschillen zijn nader geanalyseerd en bleken grotendeels verklaarbaar vanuit contextuele factoren zoals ontbrekende informatie, lokale werkafspraken of professionele inschatting.

Verklaringen voor verschillen in advies

Hoewel de AI in hoge mate richtlijnconforme en consistente adviezen genereerde, traden in sommige casussen verschillen op tussen model, richtlijn en jeugdarts. De analyse laat zien dat deze verschillen zelden wijzen op fouten van het model of van de arts, maar eerder voortkomen uit contextuele nuance. Verschillen tussen het AI-model, de richtlijn en het oordeel van de jeugdarts bleken in vrijwel alle gevallen verklaarbaar vanuit de context van het consult. Om deze context tastbaarder te maken, worden in de tabellen concrete voorbeelden gepresenteerd van situaties waarin verschillen optraden. Tabel 2 illustreert casussen rondom heupdysplasie, terwijl tabel 3 casussen met huidafwijkingen beschrijft. De tabellen tonen per casus het beleid zoals voorgesteld door de jeugdarts,

zoals verwoord in de richtlijn en zoals geadviseerd door het AI-model. De kolom 'reden van verschil' verduidelijkt waar de afwegingen uiteenlopen en waarom.

Drie hoofdoorzaken van verschil komen naar voren:

- *Ontbrekende details in het dossier.* In sommige casussen waren essentiële gegevens, zoals het verloop van klachten of eerdere observaties, niet in het dossier vastgelegd. De arts kon hierop anticiperen tijdens het consult, terwijl het model beperkte toegang had tot deze impliciete context. Dit verklaart bijvoorbeeld het verschil in casus 1 en 3 (tabel 2) en casus 6 (tabel 3).
- *Regionale werkafspraken.* Jeugdartsen opereren vaak binnen regionale netwerken waarin aanvullende afspraken gelden die afwijken van de landelijke richtlijn. Deze afspraken zijn doorgaans niet opgenomen in de input voor het model. Casus 4 en 7 in tabel 3 tonen hoe deze afspraken tot andere keuzes leiden.
- *Professionele ervaring en klinische intuïtie.* In enkele gevallen week de arts bewust af van de richtlijn, bijvoorbeeld op basis van een niet-pluisgevoel of eerdere ervaring. Deze vorm van impliciete kennis speelt een belangrijke rol in het klinisch handelen, maar is voor het model moeilijk te vatten. Casus 2 in tabel 2 en casus 5 en 8 in tabel 3 zijn hier voorbeelden van.

Door deze verschillen expliciet te maken, wordt duidelijk hoe menselijke expertise en context een aanvullende waarde vormen naast AI-ondersteuning. De tabellen helpen om de aard van deze verschillen te begrijpen en illustreren waar samenwerking tussen mens en model waardevol kan zijn.

Discussie

Deze verkennende studie laat zien dat een AI-model op basis van een Large Language Model (LLM) in staat is om binnen de jeugdgezondheidszorg (JGZ) richtlijnconforme beleidsadviezen te genereren met een hoge mate van consistentie. In 97 procent van de casussen kwam de classificatie van het model overeen met die van de richtlijn, en ook de overlap met het professionele oordeel van jeugdartsen was aanzienlijk. Daarmee bevestigt de studie dat LLMs potentieel bieden voor het toegankelijk maken van complexe, tekstuele richtlijnkennis – een uitdaging

Tabel 1. Overeenstemming tussen AI-model, richtlijn en jeugdarts per domein.

Richtlijndomein	Aantal casussen	Classificatie		Vervolgstappen	
		Overeenstemming AI ↔ Richtlijn	Overeenstemming AI ↔ Jeugdarts	Overeenstemming AI ↔ Richtlijn	Overeenstemming AI ↔ Jeugdarts
Heupdysplasie	12	100%	93%	75%	67%
Huidafwijkingen	14	100%	100%	64%	64%
Hartafwijkingen	4	100%	100%	100%	100%
Voeding & eetgedrag	6	100%	100%	100%	83%

NB: percentages betreffen volledige overeenstemming per casus.

waarvoor traditionele beslisbomen minder geschikt zijn. De bevindingen laten zien dat AI-adviezen beperkt zijn in het meenemen van contextuele factoren, tenzij deze expliciet in het dossier zijn opgenomen. Verschillen tussen het model en de arts deden zich vooral voor wanneer contextuele elementen – zoals verloop over tijd, impliciete klinische ervaring of regionale afspraken – een rol speelden in het oordeel van de arts. Dit bevestigt dat het model geen vervanging is van professionele besluitvorming, maar een hulpmiddel dat – mits gevoed met volledige en juiste informatie – bijdraagt aan richtlijnconform en consistent beleid. Daarnaast kan het model professionals ondersteunen bij het efficiënt verwerken van grote hoeveelheden richtlijninformatie. Het is daarbij cruciaal dat informatie die voor artsen vanzelfsprekend of impliciet is, expliciet wordt opgenomen in het digitale dossier of als input voor het model, om passende adviezen te genereren.

De resultaten sluiten aan bij recente literatuur over het belang van *trustworthy AI* in de zorg, waarbij niet alleen de output van het model telt, maar ook de aansluiting op werkprijktijken, autonomie en interpretatieruimte van de professional (Fjeld et al., 2020; FUTURE-AI, 2021). Een co-

learning-benadering, waarin AI en professional elkaar aanvullen, vraagt daarom om goed ontwerp van de gebruikersinterface: controleerbaar, aanpasbaar en transparant (Dellerman et al., 2021). Daarnaast is technische doorontwikkeling nodig, zoals het standaardiseren van prompts, het uitbreiden van richtlijnstructuren met vrije tekstverwerking en het integreren van lokale context in het model.

De casus binnen de JGZ maakt duidelijk dat generatieve AI een bruikbare aanvulling kan zijn op bestaande beslisondersteuning, juist in situaties waarin richtlijnen wél leidend zijn, maar vrije tekst en contextuele nuance onvermijdelijk blijven. Tegelijkertijd roept dit bredere vragen op over de rol van AI in klinisch redeneren: hoe maken we de stap van technisch correcte output naar contextueel passend beleid? En hoe zorgen we ervoor dat systemen die gericht zijn op consistentie ook ruimte houden voor professionele intuïtie en onzekerheid?

Deze studie levert een eerste empirische basis om die vragen te verkennen. Vervolgonderzoek is nodig in andere zorgdomeinen, met grotere aantallen casussen en in directe interactie met eindgebruikers. Alleen dan kan AI verantwoord worden ingebed in het socio-technisch zorgsysteem.

Tabel 2. Voorbeelden van verschillen tussen jeugdarts, richtlijn en het AI-model ten aanzien van heupdysplasie.

Casus	Beleid jeugdarts	Beleid richtlijn	Beleid model	Reden van verschil
1	Verwijzen voor beeldvormend onderzoek op 3 maanden	Ontbrekende informatie, geen eenduidig advies	Verwijzen voor beeldvormend onderzoek binnen 2 weken	AI geeft scherper advies op basis van richtlijn, maar mist context, zoals tijdsverloop
2	Verwijzen voor beeldvormend onderzoek op 3 maanden	Geen bijzonderheden	Geen bijzonderheden	Jeugdarts kijkt af op basis van ervaring of eerdere signalen buiten dossier
3	Verwijzen voor onderzoek binnen 2 weken	Ontbrekende informatie, geen eenduidig advies	Verwijzen naar orthopeed met verzoek tot snelle beoordeling	AI specificeert type specialist op basis van interpretatie richtlijntekst

Tabel 3. Voorbeelden van verschillen tussen jeugdarts, richtlijn en het AI-model ten aanzien van huidafwijking.

Casus	Beleid jeugdarts	Beleid richtlijn	Beleid model	Reden van verschil
4	Verwijzen naar huisarts, regulier	Verwijzen naar huisarts, regulier	Verwijzen naar dermatoloog, versneld (1-2 weken)	Regionale afspraken
5	Verwijzen naar kinderarts/dermatoloog, geen termijn	Verwijzen naar kinderarts/dermatoloog, spoed (1-2 dagen)	Verwijzen naar kinderarts/dermatoloog, versneld (1-2 weken)	Klinische ervaring en inschatting jeugdarts
6	Verwijzen naar kinderarts/dermatoloog, regulier	Verwijzen naar huisarts, regulier	Geen bijzonderheden	Ontbrekende details, zoals tijdsverloop of observaties
7	Verwijzen naar kinderarts/dermatoloog, regulier	Verwijzen naar huisarts, regulier	Verwijzen naar kinderarts/dermatoloog, regulier	Regionale afspraken
8	Verwijzen naar kinderarts/dermatoloog, spoed (1-2 dagen)	Verwijzen naar kinderarts/dermatoloog, spoed (1-2 dagen)	Verwijzen naar kinderarts/dermatoloog, versneld (1-2 weken)	Klinische ervaring en inschatting jeugdarts

Bijdrage aan het HF-kennisdomein

Dit artikel draagt bij aan het HF-kennisdomein door beslissondersteuning met AI te benaderen als onderdeel van een socio-technisch zorgsysteem waarin richtlijnen, digitale infrastructuur en professionele oordeelsvorming samenkomen. De studie belicht hoe een Large Language Model (LLM) functioneert binnen bestaande JGZ-processen en toont dat verschillen tussen model, richtlijn en professional vaak voortkomen uit systeemgrenzen: impliciete context, regionale afspraken of consultlogistiek. Daarmee wordt expliciet bijgedragen aan kennis over systeemaanpak op micro- en mesoniveau. Het *ontwerpergerichte karakter* blijkt uit de toepassing van AI voor twee functies: het ontsluiten van richtlijnkennis via RAG én het vertalen van ongestructureerde consult-data naar bruikbare input voor beleid. Deze ontwerpkeuzes worden geëvalueerd op hun bruikbaarheid en aansluiting bij het werkproces van professionals, met implicaties voor toekomstige AI-interfaceontwikkeling in de zorg.

De resultaten maken duidelijk dat AI-modellen *systeemprestaties*, zoals richtlijnconformiteit en consistentie, kunnen versterken, mits goed gevoed en ingebed. Tegelijkertijd tonen de casussen aan dat het *welbevinden* van professionals – zoals autonomie en vertrouwen in eigen oordeel – geborgd moet blijven bij de implementatie van AI. Daarmee biedt het artikel concrete aanknopingspunten voor het balanceren van technische systeemoptimalisatie en menselijke waarden in zorginnovatie.



Abstract

AI-based decision support systems offer opportunities for preventive child healthcare, yet often clash with real-world practice where guidelines are complex and medical records are largely unstructured. This study explores how AI-generated advice aligns with both clinical guidelines and professional judgment. Fifteen pediatric cases were collected from Youth Health Care (JGZ). For each case, a policy recommendation was formulated by a youth doctor, based on official guidelines, and by a large language model (LLM). Recommendations were compared on classification and proposed actions. In 97 percent of cases, the LLM's classification matched that of the guideline. Action recommendations matched in 78 percent of cases. Deviations were mostly explained by missing context, regional agreements, or the doctor's clinical intuition. LLMs can provide consistent, guideline-conform advice. However, human context and expertise remain essential. The model should support, not replace, professional decision-making in preventive child health care.

Referenties

- Blanson Henkemans, O.A., Werges, H., Peute, L., van den Berg, L., Bezem, J., Vlasblom, E., & Vrijmoeth, M. (2021). eHealth voor uniforme en passende jeugdgezondheidszorg: Beslissondersteuning met slimme richtlijnmodule via digitale dossiers. *Tijdschrift voor Human Factors*, 46(3), [oktober 2021]. <https://www.humanfactors.nl/tijdschrift/download/nummer-3-oktober-2021/312>
- Dellermann, D., Ebel, P., Söllner, M., & Leimeister, J.M. (2021). Hybrid intelligence. *Business & Information Systems Engineering*, 63(5), 513-523. <https://doi.org/10.1007/s12599-021-00696-z>
- Fjeld, J., Achten, N., Hilligoss, H., Nagy, A.C., & Srikumar, M. (2020). *Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI* (Berkman Klein Center Research Publication No. 2020-1). <https://cyber.harvard.edu/publication/2020/principled-ai>
- Hoekstra, M., Dom, L., & Van Veenstra, A.F. (2024). *Quickscan AI in de Publieke Dienstverlening III* (TNO-rapport 2024 R11005). TNO Vector. <https://publications.tno.nl/publication/34642601/SASnc3ZW/TNO-2024-R11005.pdf>
- Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C.H., & Kang, J. (2020). BioBERT: A pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4), 1234-1240. <https://doi.org/10.1093/bioinformatics/btz682>
- Liu, Y., Chen, P.H.C., Krause, J., & Peng, L. (2023). How to read articles that use machine learning: Users' guides to the medical literature. *JAMA*, 330(9), 844-852. <https://doi.org/10.1001/jama.2023.16255>

Over de auteurs



I. (Imme) Moeskops, MA
Arts in opleiding tot specialist (aios),
SBOH



Dr. O.A. (Olivier) Blanson Henkemans
Senior Onderzoeker, TNO
Olivier.BlansonHenkemans@TNO.nl



T. (Tim) van den Broek
Onderzoeker, TNO



A. (Arjan) Huizing
Onderzoeker, TNO