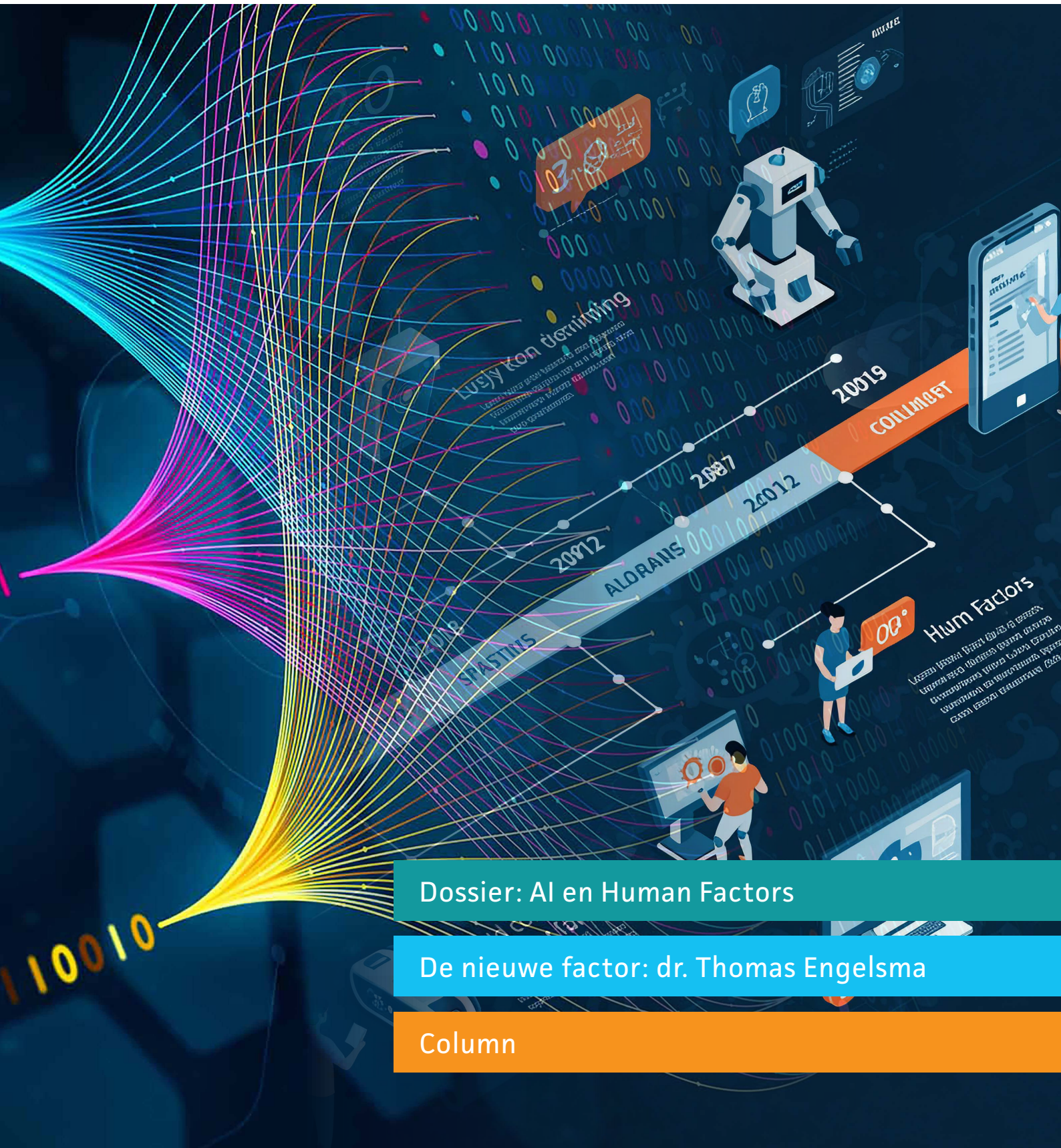
Tijdschrift voor

jaargang 50 - nr. 2 - juni 2025

HUMAN FACTORS



Dossier: AI en Human Factors

De nieuwe factor: dr. Thomas Engelsma

Column

Colofon

Human Factors streeft naar het zodanig ontwerpen van gebruiksvoorwerpen, technische systemen en taken, dat de veiligheid, de gezondheid, het comfort en het doeltreffend functioneren van mensen worden bevorderd.

Tijdschrift voor Human Factors is een uitgave van Human Factors NL, vereniging voor ergonomie. De vereniging tracht op basis van bovengenoemde omschrijving onderzoek te bevorderen, resultaten openbaar te maken, praktische toepassingen te stimuleren en uitwisseling van gegevens tussen belanghebbende vakgebieden te doen plaatsvinden.

Secretariaat van Human Factors NL
Van Nesstraat 29, 2024 DK Haarlem
leden@humanfactors.nl
www.humanfactors.nl

Redactie
dr. N.W. Wiezer, hoofdredacteur@humanfactors.nl
dr. O.A. Blanson Henkemans, olivier.blansonhenkemans@tno.nl
A. van der Have PhD, tuur.vanderhave@kuleuven.be
dr. T. Luger, tessy.luger@med.uni-tuebingen.de
ir. M. Smulders, maxim.smulders@lvnl.nl
H. Vermeulen, hettyvermeulen@vhp.nl

Redactieraad
dr. A.H.M. Cremers, prof.dr.ir. J. Dul, drs. J. Jansen, prof.dr. M.P. de Looze, dr.ir. M. Melles, prof.dr.ing. W.B. Verwey

Uitgever in opdracht
Reijseger to the point
Postbus 174, 3760 AD Soest
Telefoon: 035 693 6776
info@reijsegertothepoint.nl

Technische eindredactie
Marit Hazeleger
tvhf@reijsegertothepoint.nl

Realisatie en ontwerp
Practicum, Soest
practicum.nl

Advertenties
Advertentiewinkel.nl
Postbus 174, 3760 AD Soest
Telefoon: 035 693 6776
info@advertentiewinkel.nl

Abonnementen
Het Tijdschrift voor Human Factors verschijnt vier maal per jaar. De abonnementsprijs bedraagt € 80,- per jaar (excl. 9% btw). Abonnementen kunnen ieder moment ingaan, doch slechts worden beëindigd indien schriftelijk vóór 1 december van de lopende jaargang is opgezegd en een bevestiging daarvan is ontvangen. Bij niet tijdige opzegging wordt het abonnement automatisch met een jaar verlengd.

ISSN 2405-7924

Richtlijnen voor auteurs
Zie www.humanfactors.nl.

Persberichten
Persberichten kunt u sturen aan de redactie.

Coverfoto
Pexels

HUMAN FACTORSNL
vereniging voor ergonomie

Inhoud

Verder in dit nummer

Voorwoord

Afgelopen week was ik op vakantie op Sardinië. In een huurautootje probeerden wij met Google Maps de weg te vinden in een doolhof van hele kleine straatjes in een dorpje in de bergen. Het ging mis toen we een door Google Maps aangegeven weg inreden en we tegen het verkeer in bleken te rijden. Van de tegenligger, een lokale bewoonster die een flinke deuk aan onze vergissing overhield, begrepen we dat de verkeerssituatie recent was gewijzigd. We waren niet de enige toerist die deze vergissing maakte. Het leverde een interessante discussie op over of we Google Maps de schuld konden geven van de deuk.

Technologie, zoals AI, gaat een steeds grotere rol spelen in ons leven. Het levert kansen op (zonder Google Maps was het vinden van de weg nog veel ingewikkelder geweest), maar ook risico's (wij vertrouwen onterecht blind op de technologie). Technologie zal alleen echt effectief zijn als mensen er goed mee kunnen (samen)werken. Judith Masthoff stelde voor dit nummer een heel mooi en interessant dossier samen over de samenwerking van mensen met AI. Zoals Judith zelf in haar inleiding zegt: de artikelen in het dossier laten zien hoe belangrijk de rol is van Human Factors in het ontwerpen van deze samenwerking.

Gebruik van technologie kan een (nog grotere) uitdaging zijn voor mensen met milde cognitieve storingen of dementie. Thomas Engelsma deed onderzoek naar de barrières van deze groep voor gebruik van ondersteunende m-health-technologie. Hij ontwikkelde en evalueerde ontwerpprincipes die ontwikkelaars kunnen ondersteunen bij de ontwikkeling van m-health-toepassingen voor deze groep. In de rubriek 'De nieuwe factor' leest u een samenvatting van het proefschrift.

Daarnaast treft u in dit nummer een artikel van Erwin Speklé en Ernst Koningsveld. Zij leggen uit waarom ergonomie in de Arbowet niet als kerndiscipline is erkend, aan de hand van de geschiedenis van de ontwikkeling van deze wet. Zij schetsen de consequenties van deze situatie en vragen zich af of het niet tijd wordt voor ergonomen om te pleiten voor een formele erkenning als kerndeskundige.

Ook in dit nummer hebben we weer een column. Dit keer van Chantal Alleblas, Human Factors-specialist bij Arbo Unie. Met haar column vormt ze een mooie brug tussen het vorige dossier over 'zwaar werk' en het dossier in dit nummer.

Een hele fijne zomer en veel leesplezier gewenst.

Noortje Wiezer
Hoofdredacteur
hoofdredacteur@humanfactors.nl

De centrale rol van Human Factors in artificiële intelligentie (AI)

Traditioneel houdt Human Factors en ergonomie (HFE) zich bezig met systemen waarin mensen interactie hebben met hun omgeving. Die omgeving kan fysiek zijn, maar ook sociaal, organisatorisch of online. Het doel van HFE is om de omgeving aan de mens aan te passen, zodat het voor mensen makkelijker en prettiger wordt om hun taken goed te doen. De gestage opkomst van kunstmatige intelligentie (AI) lijkt op dit gebied veel te veranderen, met systemen die zelf beslissingen nemen.

AI is regelmatig in het nieuws, waarbij AI vaak wordt afgeschilderd als vervanger van de mens. Zelfs in de satirische tv-show van Arjen Lubach is het een terugkerend onderwerp, met het item 'AI versus Arjen'. Bij een recent familiefeestje kwam ik een oude professor tegen. Toen ik hem vertelde dat ik werk op het snijvlak van AI en Mens-Machine Interactie (HCI), reageerde hij verbaasd. Volgens hem moest ik dan recent van onderwerp zijn veranderd, want "AI is iets nieuws". In zijn beleving was AI simpelweg ChatGPT. Hij wilde me niet geloven toen ik uitlegde dat ik al meer dan dertig jaar aan AI werk en dat AI veel meer is dan grote taalmodellen. AI wordt al decennia ingezet voor talloze toepassingen, variërend van gezondheid en muziekaanbevelingen tot mobiliteit. AI kan bijvoorbeeld mensen helpen betere beslissingen te nemen of kan helpen met gigantische hoeveelheden data om te gaan zonder overspoeld te raken.

In het nieuws wordt AI afgeschilderd als iets wat grote mogelijkheden biedt, mits er verstandig mee omgegaan wordt: 'De Overheid wil dat AI in Nederland innoveert en niet discrimineert.' Maar ook als een grote bedreiging, met krantenkoppen als: 'Samenleving moet zich voorbereiden op incidenten met AI', 'Hoe gevaarlijk is AI voor de mensheid?' en 'Kunstmatige intelligentie is best bedreigend'. Er is daarnaast veel aandacht voor regulatie. Zo heeft de Europese Unie met de *AI Act* wetgeving ingevoerd die AI reguleert op basis van het verwachte risico, om tegelijkertijd innovatie te stimuleren én fundamentele rechten van burgers te beschermen. Er is ook veel aandacht voor het verbeteren van AI, met een nadruk op het voorkomen dat AI fouten maakt.

Hoewel er veel aandacht is voor het verbeteren van AI-algoritmes, wordt vaak over het hoofd gezien dat Human Factors van cruciaal belang zijn voor het ontwerpen en inzetten van AI. Naarmate systemen een meer gelijkwaardige of zelfs leidende rol krijgen, wordt het essentieel dat mensen en systemen elkaar goed begrijpen: de interactie moet natuurlijk verlopen. AI-systemen moeten mensen kunnen aanvoelen en zich automatisch aan hen aanpassen. Waar dit vroeger vooral vooraf door ontwerpers werd vastgelegd, passen systemen zich nu steeds vaker zelfstandig, *real-time* aan de gebruiker aan. Daarnaast moeten systemen in staat zijn om hun keuzes en acties uit te leggen, zodat gebruikers begrijpen wat er gebeurt en waarom. Ze moeten vertrouwen wekken – maar niet de suggestie wekken dat je ze blind kunt vertrouwen – en ruimte laten voor menselijke controle en ingrijpen waar nodig. Alleen dan kunnen mens en AI effectief en verantwoord samenwerken.

In Nederland is er veel onderzoek op het snijvlak van AI en HCI, zoals ook zichtbaar in het recente *white paper* over de stand van zaken van onderzoek en onderwijs op het gebied van Human-Computer-Interactie in Nederland (uitgevoerd in opdracht van de SIG CHI Nederland). Er is bijvoorbeeld onderzoek naar personalisatie en adaptieve systemen, interactie met chatbots en belichaamde AI (robots en virtuele mensen), en systemen die mensen coachen om gezonder en duurzamer te werken.

De drie artikelen in dit dossier geven een indruk van het type werk dat wordt verricht op het snijvlak van AI en HCI.

Dinnissen et al. belichten het Human Factors-onderzoek dat zij doen voor een populair type adaptief systeem – aanbevelingssystemen ('recommender systems') – en maken duidelijk waarom dit onderzoek van cruciaal belang is. Janssen laat zien hoe er vanuit diverse perspectieven en met verschillende methoden wordt gewerkt aan Human Factors voor AI, in een nauwe samenwerking tussen bedrijfsleven en universiteit, met mobiliteit als toepassingsgebied. Het onderzoek van Moeskops et al. naar grote taalmodellen voor beslissingsondersteuning in de gezondheidszorg laat zien dat AI contextuele factoren mist die essentieel zijn voor menselijke besluitvorming. Dit onderstreept het belang van Human Factors-onderzoek voor een effectieve integratie van AI.

Het dossier heeft een heel duidelijke conclusie: AI maakt het vakgebied Human Factors alleen maar belangrijker.

Over de gastredacteur

Prof. dr. ir. J.F.M. (Judith) Masthoff
Professor Interactie Technologie,
Universiteit Utrecht

ORGANISATIE

Aanbevelingen van een computer – hoe betrouwbaar zijn ze?

Recommender systems (letterlijk vertaald 'aanbevelingssystemen') zijn de drijvende kracht achter aanbevelingen en feeds in onder andere webwinkels, streamingdiensten en sociale media. Deze aanbevelingen hebben grote invloed op welke artikelen we lezen of kopen, of welke meningen we horen. Echter, deze aanbevelingen zijn vaak niet objectief, waardoor bestaande ongelijkheden in de samenleving kunnen vergroten. In dit artikel beschrijven we hoe transparante algoritmes en verklaringen dit tegen kunnen gaan, en presenteren we interactie-patronen die gebruikers meer controle over hun aanbevelingen geven.

Karlijn Dinnissen, Hanna Hauptmann, Eelco Herder, Judith Masthoff en Aletta Smits

In veel online platforms kunnen gebruikers kiezen uit een grote hoeveelheid zaken, zoals artikelen in een webshop, posts op sociale media, of series of muziek bij streamingdiensten. De platforms proberen hun gebruikers daarom te helpen met gepersonaliseerde aanbevelingen. We laten bijvoorbeeld streaming-diensten muziek voor ons afspelen zonder elk nummer zelf te kiezen. De sociale media stellen ook voor het grootste deel zelf onze tijdlijn samen: een door een algoritme samengestelde selectie van berichten, op basis van informatie die het platform over ons heeft.

De drijvende kracht achter aanbevelingen op het internet zijn *recommender systems* (RS). Dit zijn algoritmes die boeken, films, nieuwsartikelen en andere (online) producten voor een gebruiker selecteren. Die selectie kan plaatsvinden door het product op basis van de inhoud te koppelen aan gebruikerswensen (*content-based recommendation)* of door producten te identificeren die gebruikers met een vergelijkbare smaak hebben gekozen (*collaborative filtering*) (Ricci et al., 2010). Met name die laatste techniek wordt vaak toegepast, omdat keuzes uit het verleden een goede indicatie zijn van producten die gebruikers in de toekomst waarschijnlijk aantrekkelijk of nuttig zullen vinden.

Gebruikers zien de aanbevelingen doorgaans in de vorm van een lijst of een *feed* waaruit ze simpelweg kunnen kiezen, zoals popsongs of series die ze nog niet kennen maar wel leuk zouden kunnen vinden. Daarmee neemt het algoritme de adviesfunctie over van een winkelverkoper, of van de redacteur die een programma samenstelt. Omdat het RS-algoritme bepaalt wat in de feed komt – en wat niet – beïnvloedt het daarmee direct onze beslissingen. Hoe meer we gebruik maken van RS, des te belangrijker is het dat we goede interactiemogelijkheden hebben om deze systemen te kunnen gebruiken en controleren.

In dit artikel gaan we dieper in op de invloed van RS op ons dagelijks leven, en de invloed die wij zelf daarop kunnen uitoefenen. Dit doen we aan de hand van de volgende onderzoeksvragen, die elk in een eigen paragraaf worden behandeld.

- In welke mate beïnvloeden RS onze (dagelijkse) beslissingen?
- Hoe objectief en vrij van bias zijn RS? Wat betekent dit voor gebruikers en andere belanghebbenden van zo'n systeem?
- Kan transparantie en uitleg zorgen voor eerlijkere RS?
- Welke interactiepatronen kunnen gebruikers controle geven over RS?

Beslissingen nemen met behulp van RS

In toenemende mate vertrouwen we op apps, die veelal gebruik maken van RS, bij het nemen van dagelijkse beslissingen. Dit kunnen relatief onbelangrijke beslissingen zijn, zoals welk restaurant we die avond gaan bezoeken, maar ook onze politieke keuzes en onze nieuwsconsumptie worden deels bepaald door feeds van sociale media en gepersonaliseerde zoekresultaten (Herder et al., 2024).

In de beginjaren van het internet was het lastig om de juiste informatie of de juiste website te vinden. Dat was weliswaar frustrerend, maar juist het zoekproces en het raadplegen van verschillende bronnen en invalshoeken is nuttig om je mening te vormen. In de afgelopen twee decennia zijn zoekmachines en de algoritmes van sociale media zo volwassen geworden dat ze vrijwel altijd

relevante resultaten en suggesties laten zien, met als gevolg dat gebruikers niet verder zoeken.
Bij het bepalen wat wel of niet in de feed (of zoekresultaten) komt, treden ongewenste effecten op als filterbubbels die gebruikers in complottheorieën kunnen doen geloven, of bestaande maatschappelijke ongelijkheden versterken (Baeza-Yates, 2018), zoals de dominantie van mannelijke artiesten in de muziekwereld – meer hierover in de volgende paragraaf. Omdat RS mede bepalen welk nieuws, welke media, welke personen en welke berichten ons bereiken en welke niet, oefenen ze een macht uit die directe gevolgen heeft voor onze samenleving.
RS kunnen worden geschaard onder het (veel) bredere begrip 'artificiële intelligentie' (AI). In de media wordt doorgaans over AI (en indirect ook RS) gesproken als algoritmes die volledig zelfstandig hun gang gaan, met allerlei ongewenste gevolgen. Dit ligt echter genuanceerder: zoals we in de volgende paragraaf beschrijven, hebben de instituten die RS inzetten zelf een verantwoordelijkheid bij het vormgeven van de algoritmes. Verder – en dat is de centrale boodschap van dit artikel – kunnen en moeten de gebruikers van RS een actieve rol spelen bij het al dan niet accepteren van aanbevelingen.
Gebruikers hebben ondersteuning nodig om te bepalen of ze zoekresultaten in twijfel moeten trekken, een productsuggestie aan willen nemen, of juist om alternatieven moeten vragen. Dat is de focus van het vervolg van dit artikel, waarbij we in de komende paragraaf beginnen met *fairness* en *bias* in RS.

Eerlijkheid en objectiviteit in RS: fairness en bias

Vaak veronderstellen gebruikers dat een algoritme 'objectief' keuzes maakt, zonder inmenging van enige belangen of vooroordelen. In zekere zin is dat een logische gedachte, omdat computers inderdaad geen emotie of culturele achtergrond hebben, waar dit bij mensen wel invloed heeft op hun keuzes. Echter, algoritmes worden ontwikkeld door mensen, die ook vaak werken bij een bedrijf dat bepaalde doelen en belangen heeft die impact hebben op bijvoorbeeld wat wordt aanbevolen. Daarnaast worden algoritmes getraind op basis van data die komt uit de echte wereld, en die *bias* kan bevatten die wellicht de resultaten van het algoritme oneerlijk maakt. Zulke bias kan ontstaan op verschillende punten, waarvan we er hier enkele zullen noemen.
Ten eerste kan ongelijkheid ontstaan tijdens het verzamelen van data. Neem bijvoorbeeld een RS dat nieuwsartikelen aanbeveelt; als dat systeem uitsluitend getraind wordt op basis van wat Nederlanders die in Amsterdam wonen eerder hebben gelezen, dan is de kans groot dat aanbevelingen niet goed aansluiten bij een Engelsman die recent is geëmigreerd naar Groningen en graag op de hoogte wil raken van het nieuws in Nederland. Het systeem werkt daardoor minder goed voor de ene gebruiker dan voor de andere.
Een andere bron van *data bias* kan een ongelijke verdeling zijn die al bestaat in de 'echte wereld', en die op die manier gereflecteerd wordt in de verzamelde data. Denk bijvoorbeeld aan het feit dat technische functies historisch gezien meestal bekleed werden door mannen. Als je puur die data als uitgangspunt neemt voor een algoritme dat voorspelt welke persoon het beste past bij een technische vacature, dan zou het zomaar kunnen dat het systeem alleen mannelijke kandidaten zal aanbevelen[1]. Het is daarom belangrijk dat er wordt stilgestaan bij dit soort historische ongelijkheden, om te voorkomen dat die – impliciet of expliciet – worden overgenomen door RS.
Een verdere vorm van bias, *amplificatiebias* of ook *popularity bias*, vindt zijn oorsprong in een soort sneeuwbaleffect: als er al veel meer items worden aanbevolen van een bepaalde groep aanbieders, worden die meer geconsumeerd, waardoor ze daarna wellicht nóg vaker aanbevolen zullen worden, enzovoort. Hierdoor kunnen artiesten die al heel populair zijn, nóg populairder worden – dit ten koste van de minder bekende artiesten.
Als onderzoekers kunnen we er zelf al een behoorlijk beeld van krijgen hoe ongelijkheid in een RS geïntroduceerd, gemeten, en ook opgelost kan worden. Hieronder geven we enkele voorbeelden. Maar omdat menselijke normen en waarden zo persoonlijks en subjectief zijn, is het erg belangrijk om direct contact te zoeken met de mensen die impact ondervinden van een bepaald systeem.
Aan de Universiteit Utrecht hebben we zulk onderzoek gedaan voor RS bij muziek-streamingdiensten. Daarvoor gingen we in gesprek met artiesten om te kijken hoe groot de rol van muziek-aanbevelingen was in hun carrière, in hoeverre het voor hen duidelijk was hoe deze systemen werken, en of zij vonden dat er op het gebied van eerlijkheid voor hen als artiest nog verbeteringen te behalen waren (Dinnissen & Bauer, 2023). Op die laatste vraag waren artiesten het er unaniem over eens dat nieuwe en onbekendere artiesten een grotere kans moeten krijgen om aanbevolen te worden – en dat dus de eerdergenoemde amplificatiebias moet worden tegenwerkt.
Een ander probleem dat in de muziekwereld vaak ter sprake komt is de overrepresentatie van mannelijke artiesten. De artiesten die wij spraken waren het over het algemeen eens dat hier verandering in moest komen, maar sommigen vonden dat deze verandering niet zou moeten gebeuren door algoritmes meer muziek van niet-mannelijke artiesten te laten aanbevelen, maar eerder maatschappijbreed plaats zou moeten vinden.
Ook op het gebied van transparantie gaven zowel artiesten als anderen werkzaam in de muziekindustrie aan dat er nog veel ruimte voor verbetering is. Hieronder zetten we uiteen hoe transparantie door middel van uitleg in elkaar zit en welke principes daaraan ten grondslag liggen.

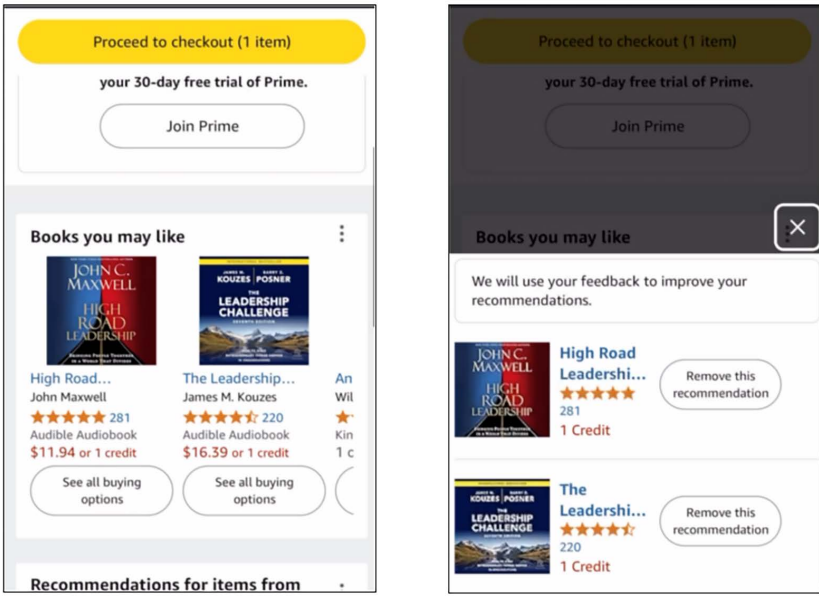
Transparantie en uitleg in RS

Om meer transparantie te bieden, wordt vaak uitleg toegevoegd aan aanbevelingen, zodat mensen begrijpen op basis van welke criteria de aanbevelingen zijn geselecteerd. Bijvoorbeeld waarom het systeem denkt dat een aanbevolen muziekstuk past bij hun smaak: “Fans van Bob Dylan zoals jij houden vaak ook van de muziek van Brandi Carlile.”

Uitleg kan echter veel meer doen dan alleen transparantie verhogen. Het kan het gemakkelijker maken om een keuze te maken, het vertrouwen in het system verhogen, tevredenheid bevorderen, gebruikers overhalen om een aanbeveling op te volgen, hen helpen om betere beslissingen te nemen, of controle geven. Uitleg kan gebruikers zelfs in staat stellen om in te grijpen als een RS slechte aanbevelingen selecteert op basis van verkeerde veronderstellingen (Tintarev & Masthoff, 2012). Uitleg kan op veel manieren gegeven worden – zie Tintarev & Masthoff, 2022. Afhankelijk van wat je wilt bereiken, heb je een ander soort uitleg nodig. Om gebruikers te helpen bij het selecteren van een film zijn tenslotte andere argumenten nodig dan bij het overtuigen dat een specifieke film voor hen de juiste keuze is. De wijze waarop uitleg gegeven wordt, moet vaak ook worden aangepast aan de persoon. Zo wensen sommige mensen meer detail dan anderen, hebben sommigen meer achtergrondkennis, of gaan meer af op de opinies van anderen. Uitleg kan dan – in plaats van aan te stippen wat mensen die op je lijken vinden, zoals hierboven – meer inhoudelijke argumenten geven: “Brandi Carlile’s verhalende trant lijkt op die van Bob Dylan.”

Deze vormen van uitleg geven de gebruikers inzicht in waarom een artiest als Brandi Carlile wordt aanbevolen, en kunnen daarmee de gebruiker in staat stellen aan te geven *waarom* ze de muziek van die artiest niet waarderen, door bijvoorbeeld de artiest zelf of juist de vermeende interesse in ‘verhalende songs’ expliciet weg te klikken.

Veel onderzoek naar uitlegmechanismen is erop gericht om het gebruikers naar de zin te maken of ze vertrouwen in het platform te geven. Daarnaast zijn commerciële platforms er zeker ook op gericht om gebruikers over halen tot consumptie. Echter, naast het ondersteunen van bestaande voorkeuren en gewoontes, kan uitleg ook een belangrijke rol spelen in het *veranderen* van menselijk gedrag, bijvoorbeeld door mensen bewust te maken van ongelijkheden in de muziekindustrie en ze te motiveren om vaker muziek van nieuwe, vrouwelijke artiesten een kans te geven.



Afbeelding 1. Interactie met applicatie voor meer controle op een recommender system. Links: controle over items in de lijst. Rechts: Verwijderen van aanbevolen items.

Interactiepatronen voor controle over RS

Een stap voorbij uitleg is gebruikers meer controle te geven over wat een algoritme aanbeveelt. Dat kan bijvoorbeeld door gebruikers te laten reageren op de aanbevelingen. Afbeelding 1 toont een voorbeeld van zo’n interactie. Als een gebruiker in deze webwinkel een boek koopt, krijgt hij vlak voor het betalen automatisch een lijst met *‘Books you may like’* te zien (linkerafbeelding). Daar hebben ze geen keuze in en dus geen controle over. Wel hebben ze een zekere mate van controle over welke boeken in de lijst komen te staan, omdat ze de lijst zelf kunnen aanpassen.

Deze website geeft de mogelijkheid om aanbevelingen uit die lijst te verwijderen (rechterfiguur). Het algoritme verzamelt daarmee informatie over de voorkeuren van de gebruiker. Bij een volgende aankoop wordt de ‘*Books you may like*’ aangepast. Voor het gevoel van de gebruiker zijn de aanbevelingen daarmee eerder een coproductie tussen algoritme en mens dan iets wat de mens door het algoritme voorgeschoteld krijgt. Deze en veel meer interacties kunnen voor gebruikers dus het gevoel van controle, transparantie en autonomie verhogen.

Over het algemeen worden er drie soorten interactie onderscheiden in de literatuur, namelijk interacties die de gebruiker helpen om de *input* waar het algoritme mee werkt aan te passen, interacties waarmee het *rekenproces* van het algoritme aangepast kan worden, en interacties die helpen om de *aanbevelingen* zelf te onderzoeken en te evalueren.

Een voorbeeld van input is een streamingdienstgebruiker die tegen het algoritme kan zeggen: ‘als je mijn *discover weekly* samenstelt, tel dan nooit de muziek mee die via mijn account tussen 6 en 8 uur wordt afgeluisterd’ (anders hebben ouders van jonge kinderen alleen Sesamstraatmuziek in hun *discover weekly* zitten). Een

voorbeeld van het aanpassen van het rekenproces is als een gebruiker tegen het algoritme van een openbaarvervoer-RS kan zeggen: 'Ik vind langere reistijd minder belangrijk dan het aantal keren overstappen.' Tenslotte is een voorbeeld van interacties die de aanbevelingen helpen evalueren de mogelijkheid om resultaten anders te sorteren, bijvoorbeeld op basis van populariteit.

Designers en data-wetenschappers ontwikkelen deze interacties gezamenlijk. Designers onderzoeken welke mate van controle de gebruiker in een specifiek RS wenst; data-wetenschappers ontwikkelen op basis daarvan een RS-algoritme dat de input van de gebruiker ontvangt en verwerkt.

Designers geven echter aan dat ze vinden dat ze te weinig van interactieve AI weten om de interacties goed te kunnen ontwerpen (Smits & van Turnhout, 2023), omdat ze vinden dat ze te weinig weten van algoritmen. Om designers te ondersteunen, ontwikkelen de Universiteit Utrecht, de Hanze Hogeschool en de Hogeschool Utrecht een bibliotheek met voorbeelden van interactiepatronen.[2]

Elke categorie interacties wordt geïllustreerd met rijke voorbeelden, die laten zien hoe ze voor de gebruiker hun gevoel voor controle, transparantie of autonomie veranderen. Er worden daarnaast ook voorbeelden gegeven van interactiemechanismen die een goed idee lijken, maar vooral een gevoel van *cognitive overload* veroorzaken: te veel knopjes waaraan gedraaid kan worden levert juist minder gevoel van controle op.

Een ander voorbeeld van minder controle is als mensen wel iets kunnen doen, maar de gevolgen ervan niet kunnen voorspellen. Neem bijvoorbeeld de *like*-button in een bekende streamingdienst: gebruikers kunnen zeggen dat ze een serie goed vonden, maar ze kunnen niet zeggen wat ze er precies goed aan vonden (afbeelding 2). Stel dat op de streamingdienst de serie *Heartstopper* een *like* krijgt omdat een acteur zo leuk is, dan is het niet meteen de bedoeling dat dit een hele reeks aanbevelingen oplevert die allemaal over het thema van die show – kalverliefde – gaan.

Afbeelding 2. Voorbeeld van een *like*-button in een applicatie waarmee invloed wordt uitgeoefend op items die worden aanbevolen.

Discussie en conclusie

In dit artikel hebben we laten zien hoe RS een essentiële rol in ons dagelijks leven spelen. RS-algoritmes zijn de drijvende kracht achter online platforms die wij gebruiken om muziek te streamen, producten te kopen, online contacten te onderhouden, of nieuws te lezen.

Hoewel wij ons graag laten leiden door wat 'in de feed' aanbevolen wordt, is het aanbevelingsproces helemaal niet zo waardevrij als veel mensen denken. Zo worden populaire keuzes en items doorgaans extra vaak aanbevolen, en ook andere bestaande ongelijkheden worden versterkt. Om deze effecten tegen te gaan, zijn transparantie en uitleg een goed begin, omdat gebruikers hiermee meer inzicht kunnen krijgen in hoe algoritmes werken en op basis waarvan aanbevelingen worden gedaan. Dit is echter niet afdoende: het is belangrijk dat de gebruiker zelf controle kan uitoefenen, zodat die zelf in kan grijpen als een RS ondermaatse aanbevelingen geeft. Daarom wordt er gebouwd aan een bibliotheek met interactiepatronen die dit mogelijk maken.

Hoe beter de interacties ontworpen zijn, des te tevredener zijn gebruikers met de aanbevolen resultaten en des te groter is de kans dat ze een platform vertrouwen – of dat nu een muziek-streamingdienst is of een persoonlijke assistent die aanbevelingen geeft over een gezondere levensstijl. En dus ook dat ze dat platform blijven gebruiken. Dat is essentieel in een wereld waarin we steeds meer adviezen uitbesteden aan AI. Want in die wereld moet de mens ook controle houden, en blijven voelen dat ze wel degelijk invloed hebben op de adviezen waar ze aan blootgesteld worden.

Abstract

Web platforms, including online stores, streaming platforms and social media, make use of recommender systems in order to create item suggestions and item feeds. These recommendations have a strong impact on which articles we read, which items we buy, or which opinions we hear. However, recommender systems are not completely neutral and may even increase existing bias within society. This article discusses how transparency and explanations can combat such unwanted effects and presents interaction mechanisms that allow users to actively manage their recommendations.

Referenties

Baeza-Yates, R. (2018). Bias on the web. *Communications of the ACM, 61*(6), 54–61. https://doi.org/10.1145/3209581

Dinnissen, K., & Bauer, C. (2023). Amplifying artists' voices: Item provider perspectives on influence and fairness of music streaming platforms. In *Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization (UMAP '23)* (pp. 238–249). ACM. https://doi.org/10.1145/3565472.3592960

Herder, E., Stojko, L., Strecker, J., Neumayr, T., Yigitbas, E., & Augstein, M. (2024). Towards new realities: Implications of personalized online layers in our daily lives. *i-com, 23*(2), 221–229. https://doi.org/10.1515/icom-2024-0017

Ricci, F., Rokach, L., & Shapira, B. (2010). Introduction to recommender systems handbook. In *Recommender Systems Handbook* (pp. 1–35). Springer. https://doi.org/10.1007/978-0-387-85820-3_1

Smits, A., & van Turnhout, K. (2023). Towards a practice-led research agenda for user interface design of recommender systems. In L. A. (Ed.), *Human-Computer Interaction – INTERACT 2023* (Vol. 14144, pp. 127–143). Springer. https://doi.org/10.1007/978-3-031-42286-7_10

Tintarev, N., & Masthoff, J. (2022). Beyond explaining single item recommendations. In *Recommender Systems Handbook* (pp. 763–796). Springer. https://doi.org/10.1007/978-1-0716-2197-4_19

Eindnoten

1 https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/

2 Dit samen met Avans Hogeschool, de Haagsche Hogeschool, de Hogeschool Rotterdam, drie internationale universiteiten, twee design-opleidingen en 15 designers.

Bijdrage aan het HF-kennisdomein

Dit artikel draagt bij aan het HF-kennisdomein door een overzicht te geven hoe recommender systems (RS) impact hebben op het gedrag en de keuzes van gebruikers. Omdat RS zijn geïntegreerd in vele online platforms die ons dagelijkse leven bepalen – zoals sociale media en streamingdiensten – is een *systeemaanpak* van belang, waarbij niet alleen de gevolgen voor individuele gebruikers worden afgewogen, maar vooral ook de effecten op de hele samenleving.

We geven een toegankelijke, systematische inleiding in de functionaliteit van RS en de oorzaken van ongewenste effecten. Als gebruikers vervolgens blind vertrouwen op de aanbevelingen die ze krijgen, vergroot dat de macht van de platforms die de keuzes aanbieden. Vervolgens wordt een *ontwerpgedreven benadering* gekozen: we geven concrete voorbeelden welke interactiepatronen transparantie bieden en de gebruikers controle geven over de werking van het aanbevelingsproces.

De rijke voorbeelden van interactiepatronen laten duidelijk zien welke invloed de *systeemprestaties* van RS hebben op het *menselijk welzijn*. Transparantie en controlemechanismen hebben een positieve invloed op de gebruikerservaring, maar vooral ook op de daadwerkelijke aanbevelingen die worden getoond en de keuzes die op basis daarvan worden gemaakt.

Over de auteurs

MSc. K. (Karlijn) Dinnissen
PhD kandidaat, Universiteit Utrecht
k.dinnissen@uu.nl

Dr. H.J. (Hanna) Hauptmann
Universitair docent, Universiteit Utrecht

Dr. Ir. E. (Eelco) Herder
Universitair hoofddocent, Universiteit Utrecht

Prof. dr. ir. J.F.M. (Judith) Masthoff
Professor Interactie Technologie, Universiteit Utrecht

Dr. A.M. (Aletta) Smits
Lector Human-Centered Technology, Hanze Hogeschool

ORGANISATIE

Mens-AI-interactie in mobiliteit

Een bloemlezing van huidig onderzoek naar menselijk gedrag in interactie met techniek in de mobiliteitssector

Binnen de mobiliteitssector zijn veel technische ontwikkelingen, bijvoorbeeld in het toepassen van kunstmatige intelligentie (AI). Er is echter ook steeds meer oog voor de mens: wat zijn de wensen en grenzen? En wanneer is AI een aanvulling op menselijk handelen? Dit artikel bespreekt een scoping review van onderzoek naar mens-AI-interactie in de mobiliteitssector, gebaseerd op verrichtingen van het ICAI AI & Mobiliteitslab, en in samenwerking met maatschappelijke partners.

Christian P. Janssen

AI en mobiliteit

Binnen de mobiliteitssector zijn er veel technische ontwikkelingen, bijvoorbeeld in het toepassen van kunstmatige intelligentie (of AI voor Artificial Intelligence). In auto's worden bijvoorbeeld sensoren geplaatst die de auto helpen goed op de weg te blijven (zoals adaptieve cruise control en laterale positieondersteuning) of om automatisch in te parkeren. Reizigersadvies-apps kunnen informatie uit meer bronnen halen om de gebruiker te ondersteunen. Infrastructuur, zoals spoorwissels en bruggen, worden met hulp van AI op afstand bediend.

In veel situaties is technologie alleen niet voldoende. Menselijke chauffeurs besturen auto's nog op verschillende soorten wegen (omdat het moet of kan) en bij het openen van bruggen doen menselijke operators het meeste uitvoerende werk. Ook als er meer automatisering komt zal er nog een rol weggelegd zijn voor mensen in veel mobiliteits- en transportsituaties. Dit is onder andere omdat diverse domeinen binnen deze sectors als 'hoog risico' zijn geclassificeerd binnen de Europese AI-act[1]. Hier is menselijk overzicht wettelijk verplicht. Wat is er dan bekend over menselijk gedrag in interactie met AI in transport en mobiliteit?

In dit artikel bespreek ik lopend onderzoek binnen het ICAI (innovatiecentrum voor AI) AI & Mobiliteitslab in Utrecht[2]. Vanuit dit lab wordt er samengewerkt tussen wetenschappers van de Universiteit Utrecht en onderzoekers uit de mobiliteitssector. Samen begeleiden zij promovendi die onderzoek doen naar een specifieke vraag die interessant is voor wetenschap en praktijk. Kernpartners van het lab zijn de NS, ProRail, KLM en QBuzz. Daarnaast zijn ook andere partners betrokken in het onderzoek. Een groot deel van het onderzoek betreft het ontwikkelen van AI-systemen en onderliggende technieken, maar ook in dit lab is er steeds meer onderzoek dat zich expliciet richt op het gedrag van de mens. De vragen die ik in dit artikel zal beantwoorden zijn:

- Welk lopend onderzoek is er naar mens-AI-interactie in de mobiliteit binnen het ICAI AI & Mobiliteitslab?
- Welke methoden worden er gebruikt in dit onderzoek?
- Welke inzichten zijn er tot nu toe uit dit onderzoek naar voren gekomen?

In de omschrijving van de projecten benoem ik steeds in welk subdomein van mobiliteit het onderzoek plaatsvindt, maar ook welke disciplinaire kennis er voor gebruikt wordt (bijvoorbeeld informatica, psychologie) en wat voor methode(s) er gebruikt worden. Deze staan ook samengevat in tabel 1.

Samenwerking tussen wetenschap en praktijk

Voordat ik op deze domein-specifieke vragen in ga, eerst een korte toelichting over wat 'samenwerking' tussen wetenschap en mobiliteitssector vanuit het lab inhoudt. In de praktijk doen promovendi onderzoek naar vragen of problemen die zijn geïdentificeerd in samenspraak tussen de partners (zoals de NS) en de wetenschappers. De promovendi werken over het algemeen ook één of meer dagen per week bij het bedrijf op locatie aan hun onderzoek. Door deze inbedding krijgen ze directer mee welke vragen er in de praktijk spelen en waar de nuances liggen die nodig zijn om onderzoek en praktijk aan elkaar te koppelen. Vaak biedt de samenwerking ook unieke kansen; bijvoorbeeld om onderzoek te doen met of voor een specifieke doelgroep (zoals operators van bruggen en sluizen) in plaats van alleen met andere studenten. De innovatie is in principe open, wat inhoudt dat de uitkomsten openbaar gepresenteerd worden. De bedrijven hebben echter het voordeel dat zij al eerder over de uitkomsten horen. Daarnaast leren verschillende partners ook van elkaar via de diverse samen-

werkingen. De verschillende bedrijven en sectoren hebben elk behoefte aan inzicht over hoe mensen omgaan met AI. Ondanks dat de specifieke vraag net iets anders is (zoals zal blijken uit onderstaande voorbeelden) zijn er ook steeds overeenkomsten waar de partners van elkaar kunnen leren. De samenwerking in het lab helpt bij deze kruisbestuiving. Omdat een promotietraject meerdere jaren loopt, is de samenwerking duurzaam. Het goed begrijpen van zowel de wetenschappelijke context als de praktijkcontext vergt enige tijdsinvestering, die minder makkelijk bij ad-hoc-samenwerkingen kan worden geleverd. De langdurige samenwerking initieert vaak ook andere projecten, zoals samenwerking op afstudeerstages van Masterstudenten.

Afbeelding 1. Set-up waarin ontvankelijkheid van het brein voor alarmen wordt getest. Foto uit Van der Heiden et al., 2022. Overgenomen met toestemming.

Onderzoek naar AI in mobiliteits- en transportsituaties

Zelfrijdende voertuigen

In mijn eigen werk heb ik onderzoek gedaan naar menselijk gedrag in automatisch rijden (wat in publieke discussies ook vaak 'autonoom rijden' wordt genoemd[3]). Het richt zich met name op de manier waarop mensen reageren op een verzoek van semiautomatisch rijdende auto's om het stuur over te nemen, bijvoorbeeld omdat de auto zelf niet weet hoe te handelen. In systeemontwerp (en onderzoek) wordt er vaak vanuit gegaan dat mensen snel reageren op dergelijke 'handover'-alarmen. Human Factors-literatuur uit andere domeinen suggereert echter dat het afhandelen van zo'n alarm niet per se instantaan is, maar dat het meer te vergelijken is met het afhandelen van een 'interruptie' – iets wat tussendoor komt terwijl je aan een taak werkt (Janssen et al., 2019). In kantoorwerk bijvoorbeeld, kunnen alarmen (zoals een 'nieuwe e-mail'-geluid) erg verstorend werken en mensen handelen dan vaak liever eerder werk nog even af tot ze een meer 'natuurlijk wisselpunt' hebben bereikt voor ze de interruptie van het alarm afhandelen.

Met deze kennis over hoe alarmen meer als een 'interruptie' benaderd kunnen worden, is het mogelijk om ook het afhandelen van interrupties in de semiautomatische auto als een proces van meerdere stadia te beschouwen, waarbij er voor elke specifieke stap (in plaats van het algemene proces) interventies kunnen worden ontworpen, gemodelleerd en getoetst (Janssen et al., 2019; Janssen, Praetorius et al., 2024). Vanuit verschillende disciplinaire invalshoeken en geassocieerde methoden hebben we inmiddels naar dit proces gekeken. In psychologieonderzoek hebben we in experimenten gemeten dat het brein wellicht minder intensief reageert op alarmen tijdens autonoom rijden (Van der Heiden et al., 2022; zie ook afbeelding 1). We hebben verschillende onderzoeken gedaan naar de interactie tussen mens en systeem, oftewel *Human-Computer Interaction*. We hebben gemeten hoe de prioriteiten van de chauffeur het tijdspad beïnvloeden waarbinnen interrupties worden afgehandeld (Ch et al., 2024). Ook hebben we gekeken naar de mate waarin het nuttig is slimme informatie (via bijvoorbeeld gekleurde waarschuwingsbalken) te delen over 'onzekerheid' die de auto heeft over de verkeerssituatie. Dit dwong de chauffeur af en toe expliciet de aandacht naar de weg te verleggen (Gerber et al., 2024). In deze onderzoeken komen methodes van experimenteren, modelleren en ontwerpen bij elkaar.

Openbaar vervoer

Meerdere promovendi van het AI & Mobiliteitslab richten zich op mens-AI-interactie in openbaar vervoersituaties. Talha Özüdoğru (zie afbeelding 2) onderzoekt samen met NS en ProRail hoe spoorwegexperts hun taken uitvoeren en hoe ze beter ondersteund kunnen worden in hun beslissingen door AI-systemen. Dit is relevant omdat de Nederlandse spoorinfrastructuur beperkt is en druk bereden wordt. Hoe drukker het spoor is, hoe meer impact een kleine verandering (zoals het weer of vertraging) kan hebben op de doorstroom van verkeer, en hoe ingewikkelder het kan zijn om een probleem op te lossen. Hoe kan vooraf een robuuste planning gemaakt worden, die al rekening houdt met alle mogelijke toekomstige onzekerheden en die ook op de dag zelf nog voldoende flexibiliteit laat om snel te kunnen handelen na incidenten, zonder dat het verkeer vertraagt? Wellicht kan AI hier in de toekomst bij helpen, door bijvoorbeeld meerdere planningen te vergelijken voor een eindkeuze, of door snel alternatieven voor te stellen bij disrupties. Via kwantitatieve experimenten probeert Talha te begrijpen: onder wat voor omstandigheden vertrouwen mensen AI-plannen wel en juist niet? En als het vertrouwen in de planning geschaad is, welke factoren beïnvloeden dan hoe snel dit hersteld wordt? In een later stadium zullen we ook de onderliggende factoren (statistisch) modelleren.

Anouk van Kasteren en Marloes Vredenborg werken ook met de NS, maar bestuderen juist de beleving van de reiziger. Hoe kan reisinformatie gepersonaliseerd worden, zodat het nog beter past bij de beleving en wensen van reizigers? De stand van zaken laat zien dat het veld nog in een vroege fase verkeert, met de meeste studies gericht op het personaliseren van route-aanbevelingen en reisinformatie, vaak met behulp van eenvoudig beschikbare contextuele gegevens, zoals locatie of tijd. Er is nog weinig aandacht voor meer geavanceerde personalisatievormen, zoals gepersonaliseerde interactie, presentatie, of het verklaren van aanbevelingen. Kansen voor toekomstig onderzoek liggen onder andere in het benutten van

minder gangbare attributen (zoals kennisniveau of sociale voorkeuren), het verbeteren van evaluatiemethoden (bijvoorbeeld met oog voor privacy, inclusie, transparantie), en het stimuleren van *open science* via betere toegang tot data en meer gestandaardiseerde rapportage (Vredenborg et al., 2025).
Daarnaast hebben ze via gebruikersonderzoek (interviews, focusgroepen) inzicht verkregen in factoren die de reiservaring beïnvloeden, zoals het weer, drukte en het tijdstip van de dag. Ze benadrukken daarbij het belang van een beter begrip van de *experienced context*, de samenhang tussen verschillende factoren, zodat reizigersinformatiesystemen beter aansluiten bij hoe mensen de wereld ervaren (Van Kasteren & Vredenborg, 2023). Naast deze contextuele factoren hebben ze onderzocht welke informatie reizigers nodig hebben tijdens verstoringen (van Kasteren et al., 2024). Op dit moment werken ze aan manieren om al deze inzichten te combineren en onderzoeken ze hoe de samenhang tussen contextuele factoren reiskeuzes beïnvloedt. Daarbij kijken ze hoe deze kennis kan worden ingezet om reisinformatie te personaliseren, zodat reizigers beter ondersteund worden bij het maken van keuzes tijdens verstoringen.

Wandelen en fietsen

Floor Bontje onderzoekt samen met TNO, uCrowds en de gemeente Utrecht digitale tweelingen van voetgangers en fietsers in stedelijke gebieden. Digitale tweelingen zijn in dit geval simulaties van verkeersstromen, die onder andere door gemeentes gebruikt kunnen worden voor het (her)inrichten van de publieke ruimte. Ze kunnen helpen bij het beantwoorden van vragen als: waar moeten blokkades of doorgangen worden geplaatst bij grote sportevenementen om de veiligheid te waarborgen? In de meeste huidige digitale tweelingen is de manier waarop het gedrag van de mens wordt gesimuleerd nog beperkt (Marçal Russo et al., 2025), wat de validiteit van de voorspellingen over menselijk gedrag ook beperkt.
Floor onderzoekt met experimenten, observaties (zie afbeelding 3) en simulaties hoe menselijk verkeersgedrag beter parametrisch gevat kan worden. Door de factoren die invloed hebben op het keuzegedrag van fietsers te kwantificeren, kunnen deze worden geïmplementeerd in digitale tweelingen. In haar onderzoek richt Floor zich in eerste instantie op verkeers- en omgevingsfactoren die de anticipatie van fietsers op voetgangers beïnvloeden. Het onderzoek toont aan dat fietsers vaker besluiten te remmen als anticipatie op een overstekende voetganger wanneer zij fietsen in gebieden met veel omringende voetgangers. Daarnaast blijkt dat fietsers hun rem-actie vaker en sneller inzetten als voetgangers dicht(er)bij zijn. Op dit moment verwerkt Floor deze data ook in computersimulaties, met cognitieve modellen, die de onderliggende factoren van keuzegedrag kunnen verhelderen, zoals eerder is gedaan voor andere mobiliteitsdomeinen, zoals autorijden (Bontje & Zgonnikov, 2024).

Afbeelding 2. Promovendus Talha Özüdoğru bij een simulatie van de werkplek van een ProRail-operator. Foto gemaakt door Chris Janssen.

Mobiliteit met een beperking

Dominique Blokland doet met Bartiméus, Robert Coppes Stichting en Visio onderzoek naar de oriëntatie en mobiliteit van mensen met visuele beperkingen. Er zijn veel verschillende vormen van visuele beperkingen. Zoals een vertroebeld zicht of een gebied (boven of onder) in het zicht niet kunnen zien. Daarnaast zijn er verschillende manieren om mensen te ondersteunen, namelijk met een geleidestok, een geleidehond en technologie, zoals sensoren, trilsignalen en gesproken aanwijzingen. Dominique brengt door middel van kwalitatieve interviews in kaart welke personen het beste gebaat zijn bij welke ondersteuning.

Nautische objecten

Rutger Stuut onderzoekt met Rijkswaterstaat hoe nautisch operators hun aandacht verdelen over meerdere schermen tijdens het op afstand bedienen van nautische objecten, zoals bruggen en sluizen. Dit kan inzicht geven voor het ontwerp van systemen en de training van medewerkers. Door middel van observaties en interviews identificeerde Rutger wat voor kijkstrategieën sluisoperators gebruiken (zie ook afbeelding 4) en hoe deze gekoppeld kunnen worden aan de taken en processen die ze doorlopen (Stuut et al., 2025). Huidig onderzoek analyseert door middel van experimenten of meer gestructureerd kijken – geleerd in een training – ook de performance van operators verbetert.

Discussie

De genoemde onderzoeken illustreren een deel van het lopende onderzoek naar mens-AI-interactie in mobiliteit, voor verschillende mobiliteitsdomeinen. Tabel 1 vat het onderzoek samen. Kenmerkend voor het onderzoek is dat er een veelzijdigheid aan methoden wordt gebruikt voor

>> Afbeelding 3. Het Janskerkhof in Utrecht, waar voetgangers die bij de bushalte uitstappen een busbaan (links) of fietspad (rechts) moeten oversteken. Floor Bontje onderzoekt het gedrag van voetgangers en fietsers op onder andere deze locatie. Foto gemaakt door Floor Bontje.

Afbeelding 4. Meting van oogbewegingen tijdens het bedienen van sluizen. Foto uit Stuut et al., 2025. Overgenomen met toestemming.

diverse domeinen. Vaak zijn meerdere perspectieven en methoden nodig om inzicht te krijgen. Kwalitatieve methoden – zoals interviews, focusgroepen en observaties – kunnen inzicht geven in de factoren die voor eindgebruikers van een technologie nuttig of wenselijk zijn, en onder welke omstandigheden. Experimenten kunnen helpen om concrete hypotheses hierover te toetsen en om de grootte van de effecten van interventies te kwantificeren op zo'n manier dat ze beter in modellen en simulaties te vatten zijn. Een dergelijke uitwisseling van ideeën kan op meerdere manieren. Het onderzoek van Rutger Stuut naar nautische objecten gaf kwalitatief inzicht in de variatie van kijkstrategieën en het opvolgende kwantitatieve onderzoek test nu hoe systematisch kijkgedrag dan daadwerkelijk is. Het onderzoek van Floor Bontje heeft juist eerst kwantitatief gemeten hoe remgedrag van fietsers afhangt van structurele factoren (zoals hoeveelheid drukte) en kijkt nu meer kwalitatief of dit zich ook voordoet in verschillende verkeerscontexten in de praktijk.

Daarnaast wordt er kennis uit diverse wetenschappelijke disciplines gebundeld om tot deze inzichten te komen. Om inzicht te krijgen in menselijk gedrag en 'performance' is er vaak een koppeling naar methoden uit de psychologie en Human Computer Interaction (HCI). Tegelijkertijd is er inzicht uit de informatica nodig over hoe de karakteristieken van AI systemen, zoals complexiteit van algoritmes in het openbaar vervoer, zijn te herleiden tot factoren die experimenteel getoetst kunnen worden. Soortgelijk zijn er relevante inzichten uit de geowetenschappen (rondom impact van de gebouwde omgeving op gedrag) en neuropsychologie (rondom cognitie en functioneren bij een beperking) nodig om het onderzoek tot nuttig inzicht te laten komen. De in dit artikel genoemde onderzoeken zijn gericht op het beter begrijpen van mens-AI-interactie en complementeert ander onderzoek in het AI & Mobiliteitslab, dat zich richt op technische innovaties, waar weer andere disciplines aan bijdragen.

Inhoudelijk laten de verschillende onderzoeken zien dat inzicht over menselijk gedrag nuttig is. Het geeft meer inzicht in de context van gebruik en helpt om kenmerken van menselijk gedrag om te zetten naar parameters die AI nodig heeft. Specifiek helpt het onderzoek om in te kunnen spelen op de wensen en verwachtingen van mensen, zoals in het onderzoek van Anouk van Kasteren en Marloes Vredenborg, voor het begrijpen van een domein voor het evalueren van trainingen, zoals in het onderzoek van Rutger Stuut en ten slotte voor het omzetten van menselijk gedrag naar parameters voor AI simulaties, zoals in het onderzoek van Floor Bontje. In elk domein zijn er steeds unieke inzichten en benodigdheden; er is (nog) niet één soort van AI die de oplossing is voor elk domein. Als lab blijven we daarom onderzoek doen in deze en andere mobiliteitsdomeinen om te zorgen dat de AI-systemen van de toekomst rekening kunnen houden met de mens – of waar we wellicht niet op AI moeten vertrouwen. Door de samenwerking van wetenschap en bedrijfsleven is er in het onderzoek een goede wisselwerking tussen praktische relevantie en fundamenteel inzicht. In de genoemde onderzoeken gaan wetenschappelijke vragen en praktijkvragen vaak hand in hand. In het nautische onderzoek kunnen bijvoorbeeld wetenschappelijke vragen getest worden over menselijke aandacht, die relevant zijn voor het ontwikkelen van specifieke trainingen voor medewerkers. Ook vormen de onderzoeken in het lab een basis voor langdurige, structurele samenwerking tussen wetenschap en praktijk. Het lab verwelkomt ideeën van samenwerking van partners; neem hiervoor contact op met Chris Janssen.

Tabel 1. Samenvatting van lopend onderzoek

Onderzoeker	Mobiliteitsdomein	Welke categorie gebruikers	Disciplines	Methode
Janssen	Autorijden	Chauffeur	Psychologie, Human Computer Interaction (HCI), Human Factors	Experimenten, modelleren, ontwerp
Özüdoğru	Openbaar vervoer	Bediener/operator	Psychologie, informatica	Experimenten, modelleren
Van Kasteren, Vredenborg	Openbaar vervoer	Reiziger	HCI, informatica	Interviews, focusgroepen, ontwerp
Bontje	Wandelen en fietsen	Reiziger	Psychologie, HCI, informatica, geowetenschap	Experimenten, observatie, modelleren
Blokland	Mobiliteit met een beperking	Reiziger	Psychologie, neuropsychologie	Interviews
Stuut	Nautisch	Bediener/operator	Psychologie, HCI, Human Factors	Observatie, interviews, experimenten

Dankwoord
Ik wil alle promovendi bedanken voor het delen van de inzichten uit hun onderzoek.

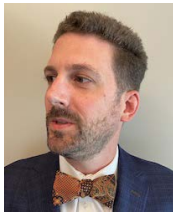
Abstract
Background. There are many technical developments within the mobility sector, for example in the application of artificial intelligence (AI). However, there is also an increasing focus of research and practice on humans: what are their needs and limitations in the use of AI? And when is AI complementary to human ability?
Method. This article provides a brief scoping review of research on human-AI interaction in the mobility sector as conducted within the ICAI AI & Mobility Lab in collaboration with societal partners.
Results. Current research on human-AI interaction focuses on different mobility domains: road, rail and water. Attention is paid to different users, both travelers and employees in the mobility sector. There is also a focus on use under special circumstances and making technology more accessible. The research is conducted using different methodologies. *Conclusion*. A good understanding of human use is an essential component for actual application of new (AI) techniques, also in the mobility sector. Cooperation between universities and social partners pays off here.

Referenties
Bontje, F., & Zgonnikov, A. (2024). How Sure is the Driver? Modelling Drivers' Confidence in Left-Turn Gap Acceptance Decisions. *Computational Brain and Behavior, 7*(3), 437-456. https://doi.org/10.1007/S42113-024-00207-7/TABLES/5
Ch, N.A.N., Fortier, J., Janssen, C.P., Shaer, O., Mills, C., & Kun, A.L. (2024). Text a Bit Longer or Drive Now? Resuming Driving after Texting in Conditionally Automated Cars. Proceedings of Automotive User Interfaces and Interactive Vehicular Applications.
Gerber, M.A., Schroeter, R., Johnson, D., Janssen, C.P., Rakotonirainy, A., Kuo, J., & Lenne, M.G. (2024). An Eye Gaze Heatmap Analysis of Uncertainty Head-Up Display Designs for Conditional Automated Driving. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. https://doi.org/10.1145/3613904.3642219
Janssen, C.P., Iqbal, S.T., Kun, A.L., & Donker, S.F. (2019). Interrupted by my car? Implications of interruption and interleaving research for automated vehicles. *International Journal of Human-Computer Studies, 130*. https://doi.org/10.1016/j.ijhcs.2019.07.004
Janssen, C.P., Praetorius, L., & Borst, J.P. (2024). PREDICTOR: A tool to predict the timing of the take-over response process in semi-automated driving. *Transportation Research Interdisciplinary Perspectives, 27*. https://doi.org/10.1016/j.trip.2024.101192
Marçal Russo, L., Dane, G., Helbich, M., Ligtenberg, A., Filomena, G., Janssen, C.P., Koeva, M., Nourian, P., Patuano, A., Raposo, P., Thompson, K., Yang, S., & Verstegen, J.A. (2025). Do urban digital twins need agents? *Environment and Planning B: Urban Analytics and City Science*. https://doi.org/10.1177/23998083251317666
Stuut, R., Janssen, C.P., Van der Stigchel, S., & Van Doorn, E. (2025). Lock Operations Through the Operator's Eyes: A Qualitative Exploration of Gaze Strategies. *Journal of Cognitive Engineering and Decision Making, 19*(1), 54-75. https://doi.org/10.1177/15553434241291260
Van der Heiden, R.M.A., Kenemans, J.L., Donker, S.F., & Janssen, C.P. (2022). The Effect of Cognitive Load on Auditory Susceptibility During Automated Driving. *Human Factors, 64*(7), 1195-1209. https://doi.org/10.1177/0018720821998850
Van Kasteren, A., & Vredenborg, M. (2023). Let's Talk About The Experienced Context: An Example Regarding Public Transport Information Systems. Adjunct Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization, 187-190. https://doi.org/10.1145/3563359.3596979
van Kasteren, A., Vredenborg, M., & Masthoff, J. (2024). Understanding Commuter Information Needs and Desires in Public Transport: A Comparative Analysis of Stated and Revealed Preferences. International Conference on Human-Computer Interaction, 14733 LNCS, 83-103. https://doi.org/10.1007/978-3-031-60480-5_5/TABLES/4
Vredenborg, M., van Kasteren, A., & Masthoff, J. (2025). Personalization in Public Transport Passenger Information Systems: A Systematic Review and Framework. *ACM Computing Surveys*. https://doi.org/10.1145/3721478

Eindnoten
1 https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai
2 https://www.uu.nl/onderzoek/ai-labs/ai-mobility-lab
3 De term 'autonoom rijden' kan impliceren dat een autonome auto eigenschappen heeft die tot volledig zelfstandig – autonoom – handelen leiden (inclusief bijvoorbeeld eigen intenties), zonder enige interactie met mensen. In de praktijk is de grote meerderheid van de huidige 'geautomatiseerde' auto's niet in staat tot dergelijk volledig zelfstandig handelen. De term 'automatisch' dekt daarom een grotere groep voertuigen.

Bijdrage aan het HF-kennisdomein
In de diverse onderzoeken wordt gekeken naar menselijk gedrag binnen een groter (verkeers)systeem. In de meeste onderzoeken ligt de nadruk op het microniveau. Er wordt gekeken naar hoe menselijk gedrag op dit moment is óf in de toekomst er uit kan zien, wat kan leiden tot advies over (her)ontwerp. Er is hierin aandacht voor zowel *systeemprestaties* (in het bijzonder veiligheid in het verkeer) en menselijk *welbevinden* (namelijk: inzet bij zinnig werk, ondersteuning en expansie van capabiliteit door techniek).

Over de auteur



Dr. C.P. (Christian) Janssen
Universitair hoofddocent psychologie van de mens-AI-interactie
AI & MobiliteitsLab & Experimentele Psychologie
Universiteit Utrecht
c.p.janssen@uu.nl

ORGANISATIE

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 ontlenen 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 tekstinvoer 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?

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.

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

Casusverzameling en classificatie

Er werd gewerkt met vijftien kind-casussen. Negen hiervan zijn verzameld bij artsen in opleiding tot specialist (aiossen), 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.

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:

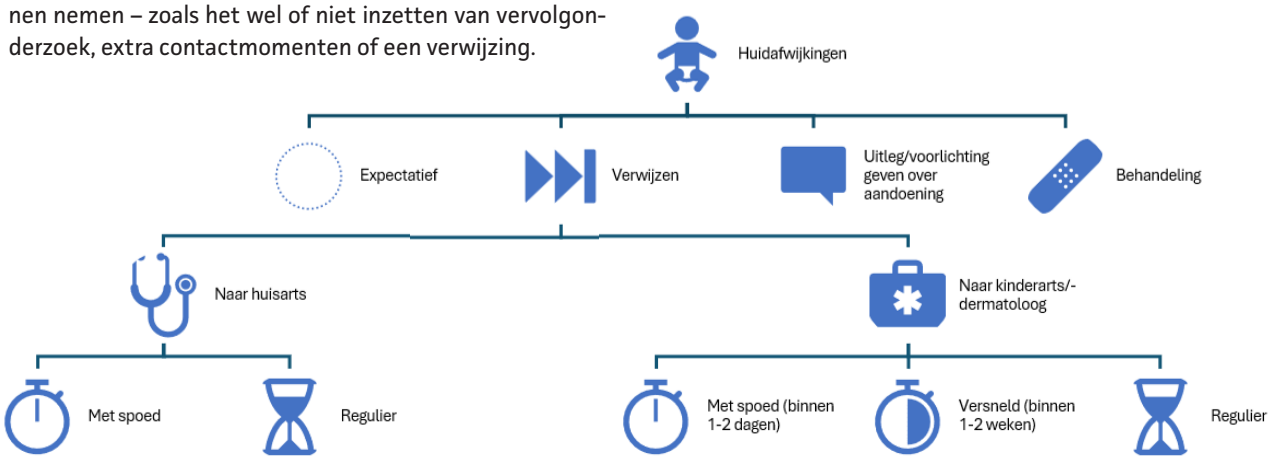
- professioneel advies: het originele beleid zoals vastgesteld door de jeugdarts;
- richtlijnadvies: het advies op basis van de geldende JGZ-richtlijn;
- AI-advies: output van de LLM.

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.

		Classificatie		Vervolgstappen	
Richtlijndomein	**Aantal casussen**	**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 ver-werken van grote hoeveelheden richtlijninformatie. Het is daarbij cruciaal dat informatie die voor artsen vanzelfsprekend of impliciet is, expliciet wordt opge-nomen 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 werkpraktijken, 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 wijkt 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 beslisondersteuning 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 *ontwerpgerichte karakter* blijkt uit de toepassing van AI voor twee functies: het ontsluiten van richtlijnkennis via RAG én het vertalen van ongestructureerde consultdata 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: Beslissingsondersteuning 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

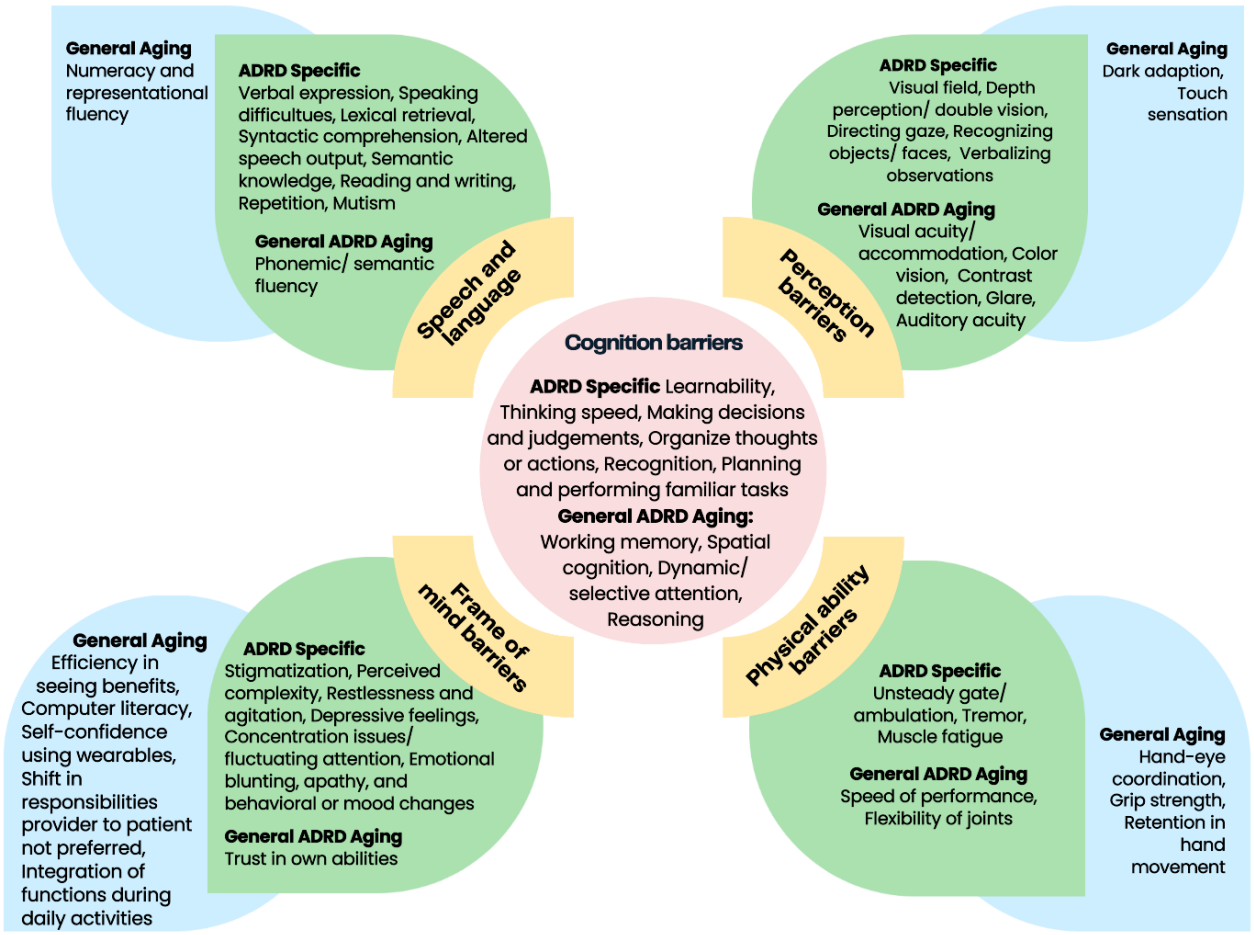
Naam: **dr. Thomas Engelsma**
Promotie: Amsterdam UMC/Universiteit van Amsterdam, eHealth Living & Learning Lab Amsterdam
Promotor: prof. dr. Monique Jaspers
Co-promotor: dr. Linda Dusseljee-Peute
Periode: 2019-2024
Titel proefschrift: Mind the Design: Optimizing the design of digital health technologies for people living with dementia

Zien wat belemmert, ontwerpen wat helpt: naar gebruiksvriendelijke technologie voor mensen met dementie

Steeds meer mensen leven met dementie en dat roept een belangrijke vraag op: hoe kunnen we zorgen dat zij zich ondersteund voelen, in contact blijven met anderen en zo zelfstandig mogelijk hun leven kunnen leiden? Digitale gezondheidstechnologie, vanaf nu gewoon 'technologie', kan daar een waardevolle rol in spelen. Denk aan apps die helpen herinneren, draagbare sensoren, videobellen met zorgverleners, platforms voor sociale interactie of hulpmiddelen voor mentale stimulatie. Deze toepassingen beloven niet alleen praktische ondersteuning, maar ook meer regie, verbondenheid en levenskwaliteit voor zowel mensen met dementie als hun naasten. Echter, de grote variabiliteit in symptomen van dementie vormt grote uitdagingen voor het ontwerpen van technologie op een manier dat deze gebruiksvriendelijk is voor deze doelgroep. Door deze verschillen zien we dat wat voor de één wel werkt, niet per se werkt voor de ander. Om dit verder te onderzoeken, heb ik mij in mijn proefschrift gericht op het systematisch benaderen van deze uitdaging. Het doel: potentiële barrières in het gebruik van technologie, specifiek mobiele gezondheidstechnologie (mHealth) voor mensen met dementie, identificeren en vertalen naar concrete, evidence-based ontwerpprincipes en richtlijnen.

Ontwikkelen van het MOLDEM-US-raamwerk

Aan het begin van mijn promotieonderzoek hebben we een scoping review uitgevoerd naar barrières in het gebruik van mHealth door mensen met milde cognitieve stoornissen (MCI) of dementie. Deze studie combineert wetenschappelijke gebruiksvriendelijkheidstudies met klinische literatuur uit het dementieveld, waar werd gekeken naar dementie-gerelateerde klachten die de gebruiksvriendelijkheid van individuele technologieën (aan de hand van casestudies) negatief hebben beïnvloed. De resultaten hiervan zijn geanalyseerd en waar van toepassing toegevoegd aan het al bestaande MOLD-US-raamwerk, ofwel *MHealth for OLDer adults* – Usability. Dit raamwerk is oorspronkelijk ontworpen om barrières voor het gebruik van mHealth voor ouderen in kaart te brengen. MOLD-US is dus uitgebreid tot MOLDEM-US, ofwel *MHealth for OLder adults with DEMentia – USability*. In dit raamwerk staan 53 potentiële barrières voor het gebruik van mHealth, ervaren door ouderen met MCI of dementie, gecategoriseerd in vijf hoofddomeinen: cognitie, waarneming, fysieke vermogens, gemoedstoestand, en spraak & taal (afbeelding 1). Dit uitgebreide raamwerk biedt een



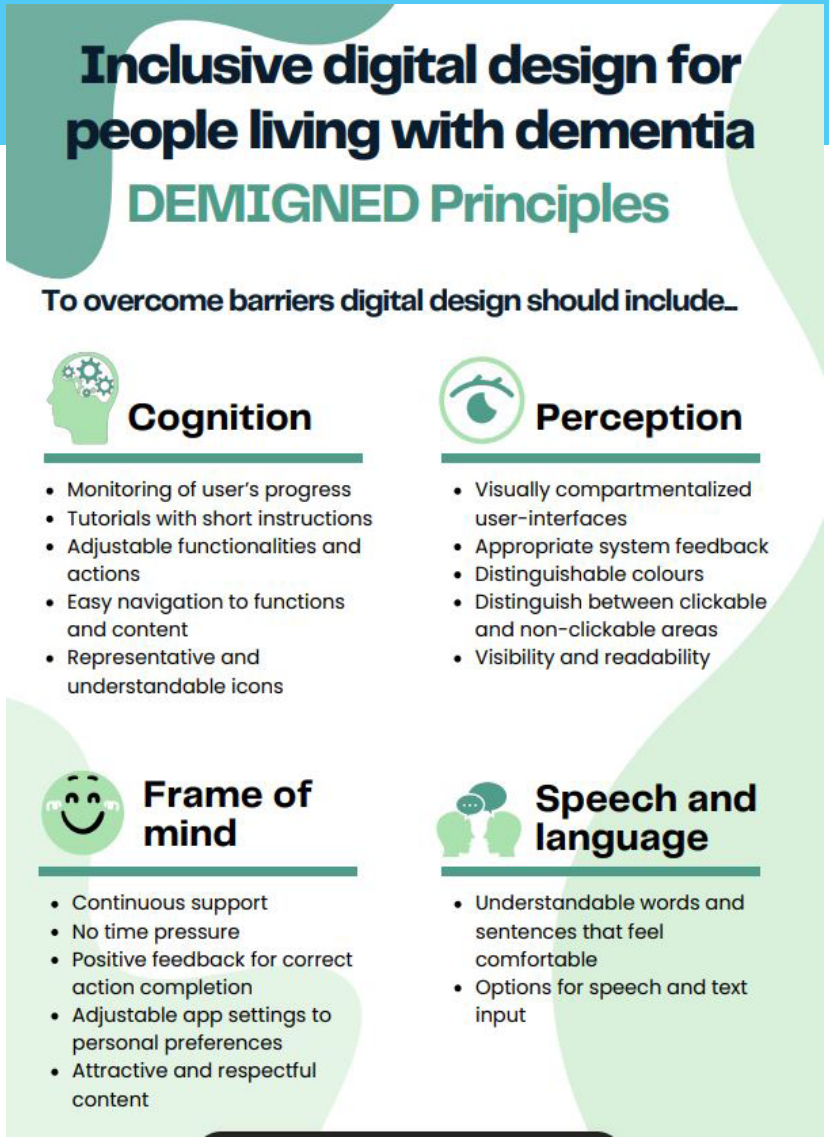
Afbeelding 1. MOLDEM-US-raamwerk.

theoretisch gefundeerd overzicht van de uitdagingen die mensen met dementie (kunnen) ervaren tijdens het gebruik van mHealth en vormt de basis voor het verdere onderzoek in mijn proefschrift.

Expert input op het MOLDEM-US-raamwerk

Omdat de barrières in het MOLDEM-US-raamwerk gebaseerd zijn op individuele studies waar mensen met MCI of dementie aan hebben deelgenomen, wilden we op hoger niveau inzichten verkrijgen in de ervaringen van experts. Zij zien immers in hun dagelijkse praktijk meerdere mensen die leven met dementie en kunnen geïnformeerde uitspraken doen over hoe symptomen van dementie kunnen leiden tot barrières in het gebruik van technologie en hoe dit kan verschillen per persoon. Ongeveer vijftig experts uit verschillende domeinen, waaronder geriatrie, neurowetenschap, thuiszorg en (wijk)verpleegkunde, werd gevraagd wat zijn zien als barrières tot het gebruik van technologie voor mensen met dementie. Dit werd hen gevraagd zonder dat ze het MOLDEM-US-raamwerk hadden gezien. De kwalitatieve bevindingen werden naast het MOLDEM-US-raamwerk gelegd om de volledigheid en relevantie van het raamwerk te beoordeelden. De barrières genoemd door de experts werden allemaal teruggevonden in het MOLDEM-US-raamwerk en bieden daarmee een eerste bevestiging van de volledigheid van het raamwerk. Naast een kwalitatieve vragenlijst werden de barrières in het MOLDEM-US-raamwerk met behulp van een online Delphistudie gekwantificeerd op impact en frequentie, waarbij experts op basis van hun praktijkervaring beoordeelden in hoeverre elke barrière het gebruik van mHealth beïnvloedt en hoe vaak zij dit zien voorkomen binnen de doelgroep. Na twee Delphi-rondes werd consensus bereikt. Uiteindelijk werden 26 barrières geïdentificeerd die zowel de hoogste impact hadden als het meest fre-

quent voorkwamen, met name binnen het domein 'gemoedstoestand' (frame of mind). Voorbeelden hiervan zijn de ervaren complexiteit, moeite met het inzien van voordelen, verminderd vertrouwen in eigen kunnen, onrust en agitatie, en moeite met concentratie. Ook cognitieve aspecten werden gezien als hoog impactvol en frequent, zoals moeite met het leren van nieuwe dingen, een verminderd werkgeheugen en redeneervermogen, en problemen met het ordenen van gedachten, plannen en het uitvoeren van vertrouwde taken. Deze inzichten helpen onderzoekers en ontwikkelaars om scherp te krijgen welke ontwerpuitdagingen echt om actie vragen. Zo kunnen keuzes in het ontwerpproces bewuster en gerichter worden gemaakt.



Afbeelding 2. Overzicht DEMIGNED richtlijnen.

Ontwikkeling van de DEMIGNED-ontwerpprincipes

Om ontwerpprincipes op te zetten voor het ontwikkelen van technologie voor mensen die leven met dementie is gestart met een verkennende literatuurstudie waarin ontwerpkeuzes uit eerder onderzoek worden gebundeld die de gebruiksvriendelijkheid van mHealth-toepassingen voor mensen met milde cognitieve achteruitgang (MCI) of dementie bevorderen. Door thematische analyses hebben we deze keuzes kunnen linken aan vier categorieën uit het MOLDEM-US-raamwerk: cognitie, perceptie, gemoedstoestand en spraak en taal. We vonden zeventien terugkerende thema's. Binnen de categorie cognitie bleek het belangrijk om geheugensteunen te bieden, duidelijke uitleg te geven, personalisatie mogelijk te maken en consistente navigatiestructuren te hanteren. Op het gebied van perceptie werd het belang onderstreept van gestructureerde scherminhoud, continue zichtbaarheid van systeemfeedback, kleurgebruik en visuele of auditieve steuntjes. Wat betreft gemoedstoestand droegen gemakkelijke toegang tot ondersteuning, tijdsoriëntatie, positieve feedback, personalisatie en anti-stigmatiserend ontwerp bij aan betere gebruiksvriendelijkheid. Tot slot bleek in de categorie spraak en taal het gebruik van eenvoudige taal, expliciete terminologie en ondersteuning voor tekst-naar-spraak- en spraak-naar-tekst-functies bij te dragen aan gebruikersgemak. Door voort te bouwen op deze thematische basis zijn de terugkerende thema's geformaliseerd in de DEMIGNED-principes (afbeelding 2). De naam DEMIGNED komt voort uit de termen: 'designed', 'mind' en 'dementia'. Elk principe is vergezeld door praktische richtlijnen voor implementatie, allen voortkomend uit de wetenschappelijke literatuur. Afbeelding 3 laat zien hoe we dit hebben gedaan voor de categorie 'frame of mind' (gemoedstoestand). Dit helpt niet alleen duidelijk te maken wat ontwerpers en ontwikkelaars kunnen nastreven, maar ook hoe deze op concrete manieren bereikt kunnen worden.

Gebruik van DEMIGNED in de praktijk

Om het gebruik van de DEMIGNED-principes te evalueren, hebben we ze toegepast tijdens de evaluatie van een informatieve mobiele website voor bezoekers van

Inclusive digital design for people living with dementia

DEMIGNED principles and actionable guidelines

Frame of mind

Continuous support

- Ensure continuous support, for instance through a helpdesk. This can be initiated through the action bar or a panic button
- Implement mechanisms to recover from errors smoothly, such as an undo button

No time pressure

- Allow ample time to respond or react to improve acceptability and prevent stress
- Timers should count up instead of down
- Provide orientation to time
- Implement time-based triggers to ensure recognition of tasks that should be completed in a recurrent interval

Positive feedback for correct action completion

- Provide failure-free content
- Allow to set goals and receive rewards to increase feelings of success
- Confirm correct steps to complete an action
- Provide brief encouragements during task completion

Adjustable app settings to personal preferences

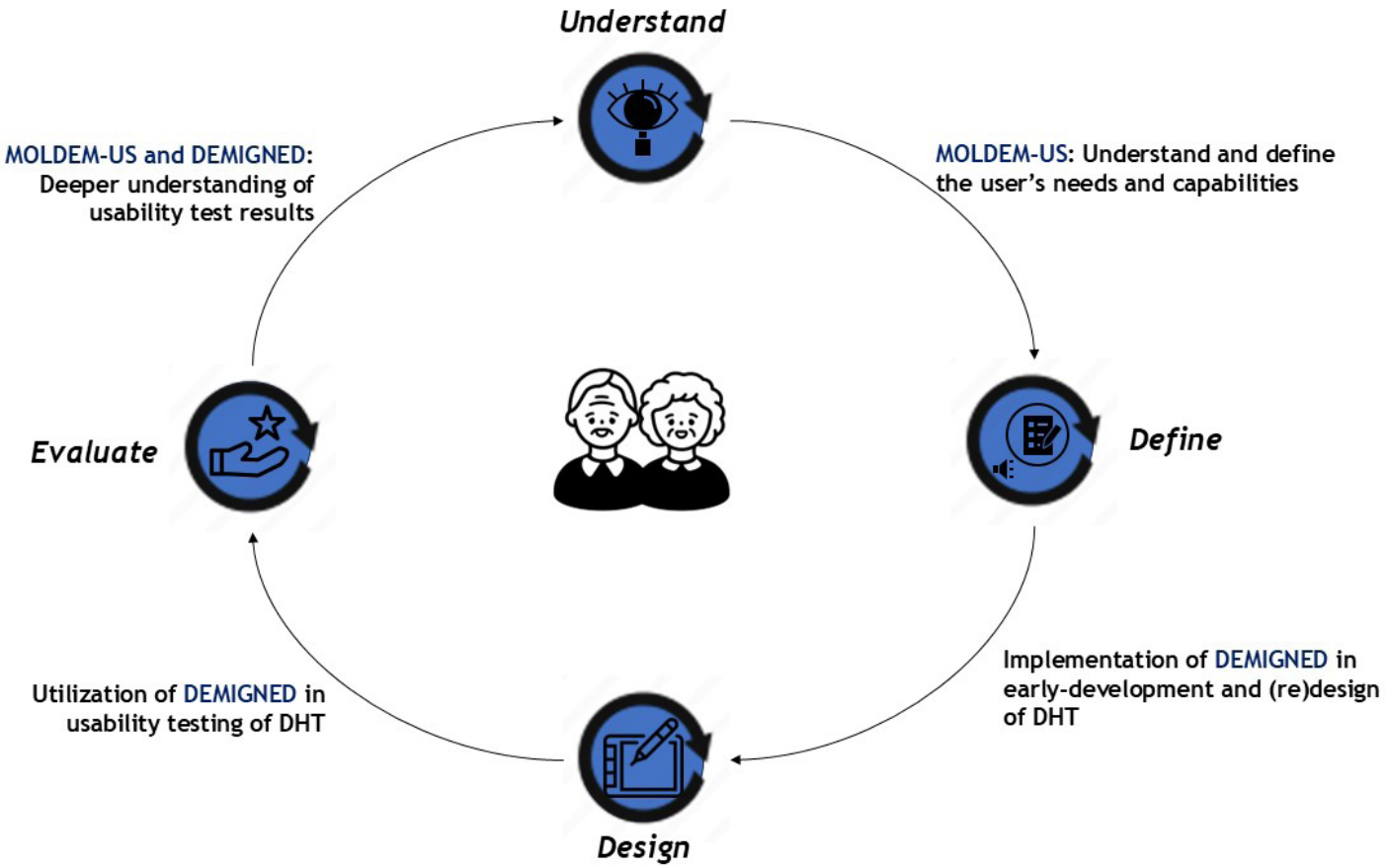
- Pre-set difficulty levels of functionalities based on educational and cognitive level
- Adjust or customize features to individual's needs and wishes
- Implement functionality adaptations as a series of options to enhance accessibility
- Implement personalized privacy settings

Attractive and respectful content

- Language should be appropriate, taking into account age- and (health) literacy-appropriateness
- Prevent stigmatization
- Use colours in the interface design and graphics to attract interest

Afbeelding 3. Voorbeeld DEMIGNED Frame of Mind (FoM).

een geheugenpoli. Deze studie is uitgevoerd met behulp van twee usability evaluatiemethoden: een expert usability studie in de vorm van een heuristische evaluatie, en een gebruikersevaluatie via 'think-aloud'-sessies met daadwerkelijke bezoekers van de geheugenpoli. Deze groep werd gerepresenteerd door mensen met subjectieve cognitieve achteruitgang, MCI of een beginnende vorm van dementie. Het doel van deze studie was om te vergelijken welke potentiële gebruiksvriendelijkheidsproblemen experts konden identificeren door de interface van de website te analyseren aan de hand van de DEMIGNED-richtlijnen, met de daadwerkelijk ervaren problemen tijdens het gebruik van de website door bezoekers van de geheugenpoli. Uit de analyses bleek dat 50 procent van de gebruiksvriendelijkheidsproblemen die experts identificeerden met behulp van de DEMIGNED-principes, ook daadwerkelijk werden ervaren door gebruikers. Omgekeerd werd 80 procent van de keren dat een probleem werd ondervonden door een bezoeker van de geheugenpoli, eveneens herkend door experts tijdens de heuristische evaluatie. Dit geeft aan dat DEMIGNED niet alleen theoretisch onderbouwd is, maar ook daadwerkelijk inzicht geeft in praktische gebruiksproblemen. Daarnaast hopen we dat de resultaten van studies zoals deze worden ingezet voor het verder verfijnen van de DEMIGNED-principes en richtlijnen. Omdat de richtlijnen voortkomen uit bestaande kennis, is het belangrijk dat ze regelmatig worden bijgewerkt op basis van nieuwe inzichten. Tijdens de studie in mijn proefschrift kwam al een belangrijke bevinding naar voren: het gebruik van de zoekfunctie op



Afbeelding 4. De (vereenvoudigde) fasen van mensgericht ontwerpen: begrijpen, definiëren, ontwerpen en evalueren.

de website sloot niet altijd aan bij het mentale model van de gebruiker. Wanneer die aansluiting ontbreekt, ondervinden mensen moeite met het vinden van informatie, terwijl dit juist hun voorkeursstrategie bleek om informatie te vinden. Dit was nog niet verwerkt in een van de richtlijnen en draagt daarmee dus bij aan de verdere verfijning ervan.

Conclusie

Het MOLDEM-US-raamwerk en de DEMIGNED-principes vormen praktische, evidence-based hulpmiddelen om (onderzoek naar) het ontwerp en de evaluatie van technologie, zoals mHealth, voor mensen die leven met dementie te verbeteren. Beide hulpmiddelen kunnen worden toegepast binnen de (vereenvoudigde) fasen van mensgericht ontwerpen: begrijpen, definiëren, ontwerpen en evalueren (afbeelding 4).

Mensgericht ontwerpen draait om het verdiepen in de behoeften, gedragingen en ervaringen van een groep eindgebruikers, om van daaruit helder te formuleren welk probleem moet worden aangepakt. Het MOLDEM-US-raamwerk ondersteunt in de vroege fasen van dit proces door systematisch in kaart te brengen welke barrières mensen met dementie kunnen ervaren. Door gebruikerseisen te koppelen aan deze barrières, krijgen ontwerpers, onderzoekers en ontwikkelaars beter inzicht in waarom bepaalde knelpunten ontstaan en welke gebruikers daar het meest door worden geraakt. Dit maakt het mogelijk om ontwerpkeuzes gerichter te prioriteren. Daarnaast kan het raamwerk ook gebruikt worden om data uit usability studies te analyseren en een meer diepgaand begrip te krijgen waarom een bepaald probleem zich voordoet. In de ontwerp- en evaluatiefasen bieden de DEMIGNED-principes een praktische vertaling van toepasbare richtlijnen naar het inclusief ontwerp. Ze kunnen ondersteunen bij het sturen van beslissingen rond interactie en interfaceontwerp, het vroegtijdig signaleren van potentiële gebruiksvriendelijkheidsproblemen met behulp van heuristische evaluaties, en (net zoals MOLDEM-US) ondersteunend zijn tijdens het analyseren van usability data. Zo ondersteunen ze het volledige iteratieve proces van HCD, waarbij continue evaluatie met eindgebruikers nog steeds centraal staat om te borgen dat oplossingen daadwerkelijk aansluiten bij hun wensen en mogelijkheden.

Met dit onderzoek hopen we bij te dragen aan de ontwikkeling van warme technologie: digitale oplossingen die niet alleen functioneel zijn, maar écht aansluiten bij het leven, de behoeften en waardigheid van mensen die leven met dementie. Technologie moet ondersteunen zonder te overheersen, verbinden in plaats van vervangen. Door barrières inzichtelijk te maken en ontwerpprincipes te bieden, wil ik met mijn proefschrift een stap zetten richting betekenisvolle en mensgerichte innovaties.

De transitie van fysiek zwaar werk

Chantal Alleblas

Werk in de sectoren logistiek, bouw, techniek of zorg is anno 2025 nog vaak fysiek te zwaar. Mensen die hier al decennia werkzaam zijn, zeggen dan: 'ach, dat hoort er nu eenmaal bij', 'ik hoef straks niet meer naar de sportschool' en 'toen wij hier net begonnen, was het allemaal nog veel zwaarder'. Aan de andere kant zijn er ook mensen die klachten ervaren en spreken over de noodzaak van 'ontziemaatregelen' en 'aangepast werk'. Dit lijkt misschien een wat gechargeerd beeld, maar toch hoor je deze twee kanten vaker onder de populatie harde werkers dan: 'het wordt tijd dat hier taken worden geautomatiseerd' of 'ons werk kan beter worden met robots en kunstmatige intelligentie, dat zou mijn werk ook leuker maken'.

Waarom zijn technologische innovaties en automatisering nog zo'n beperkt onderdeel van de belevingswereld van veel werknemers in zware beroepen? Dit in tegenstelling tot het management van organisaties, waar het vaak wel over de inzet van technologie gaat, zoals automatisch geleide voertuigen of het inzetten van (collaboratieve) robots. Meestal in relatie tot productiviteit en duurzame inzetbaarheidsproblematiek. Is het een machocultuur en angst voor het verlies van banen, of wordt er gewoon nog te weinig uiting gegeven aan de (potentiële) voordelen van samenwerken met nieuwe technologie? Hoewel de implementatie en acceptatie van technische middelen in het verleden vaak niet slaagden, is dat nu in toenemende mate wel het geval. Dit komt vooral door vooruitstrevende organisaties die kiezen voor een participatieve aanpak bij het oplossen van knelpunten, verbeteringen in onderzoek en ontwikkeling richting praktische toepasbaarheid en een steeds beter beeld van de haalbaarheid en schaalbaarheid van technische innovaties.

Het tempo waarmee innovaties worden ontwikkeld en de transitie of het verandertraject naar automatisering en robotisering varieert per sector en bedrijf. In sectoren als de logistiek en land- en tuinbouw verloopt de transitie al enige tijd relatief snel. In andere sectoren, zoals de bouw en de zorg, verloopt de transitie langzamer, onder andere vanwege de complexiteit van de werkomgevingen.

Door repetitieve en zware taken te automatiseren, kunnen bedrijven niet alleen hun productie verhogen of op peil houden, maar ook het risico op klachten aan het houdings- en bewegingsapparaat en andere gezondheidsproblemen verminderen. Een positieve ontwikkeling, maar niet zonder het risico nieuwe knelpunten te introduceren. Sterker nog, dit risico kan toenemen en daar moeten we alert op zijn. Welke functie-inhoud blijft er over, is dat rechtvaardig en bovendien, in relatie tot de toenemende krapte op de arbeidsmarkt, ook aantrekkelijk?

Om te begrijpen wat er in een arbeidsproces ondersteund kan worden met geavanceerde technologieën is een holistisch beeld van het gehele arbeidsproces en de keten nodig. De nieuwe uitdagingen binnen de kaders van industrie 5.0 vragen wellicht om een vernieuwde visie en strategie vanuit ons vakgebied. De recent door de arbeidsinspectie onder de loep genomen organisaties, zoals bagage-afhandelaren en pakketdiensten, zijn dermate omvangrijk en complex dat het risico op blinde vlekken en nieuwe knelpunten gedurende transities of verandertrajecten groot is. Bevindingen uit de arbeidsinspecties geven aanleiding voor acute interventies (veelal hulpmiddelen, met de nodige implicaties) in aanloop naar de introductie van geavanceerde systemen. Parallel zouden organisaties zichzelf ten doel moeten stellen om een concrete visie op de toekomst van werk te genereren.

Dit pleit voor een sterkere uitwerking van de 'T' in onze TOP-strategie, maar natuurlijk niet zonder 'O' en 'P'. Daarbij kunnen we ons onder meer de volgende vragen stellen: 'Welke taken kunnen het beste geautomatiseerd worden volgens de huidige stand der techniek?' 'Wat kunnen robots goed doen en wat nog niet?' 'Waar liggen de uitdagingen en waar zijn we nabij een doorbraak?' 'Kunnen technologische innovaties succesvol zijn zonder niet-technologische innovaties? Zo niet, wat is daar dan nodig?' 'Hoe ver gaan we in het uitrollen van de participatieve aanpak om de visie en ervaring van de operatie mee te nemen?' 'Hoe laat je de harde werkers voldoende deel uitmaken van een verandertraject of transitie?'

Gefaseerde ontwikkeling richting een nieuwe standaard omtrent (nu nog) zware arbeid vraagt om een bijzondere samenwerking en betrokkenheid van kennisdomeinen in verbinding met afgevaardigden uit de operatie en belanghebbenden van de organisatie(s) waar het probleem zich afspeelt. Deze samenwerkingen en connecties zijn essentieel voor het ontwikkelen van holistische en gedragen oplossingen binnen complexe ecosystemen, zoals de bagageafhandelaren en pakketdiensten.

De betrokkenheid van kennisdomeinen is ons natuurlijk niet onbekend. Juist omdat de ergonomiediscipline een diversiteit aan domeinen kent, te weten cognitieve, sensorische, organisatorische, fysieke en gedragsergonomie, wordt met de toenemende complexiteit omtrent het automatiseren van fysiek zwaar werk ook een steeds sterker beroep gedaan op de brede kijk op arbeidsprocessen en werksituaties van de ergonoom.

Moedig voorwaarts!. ■

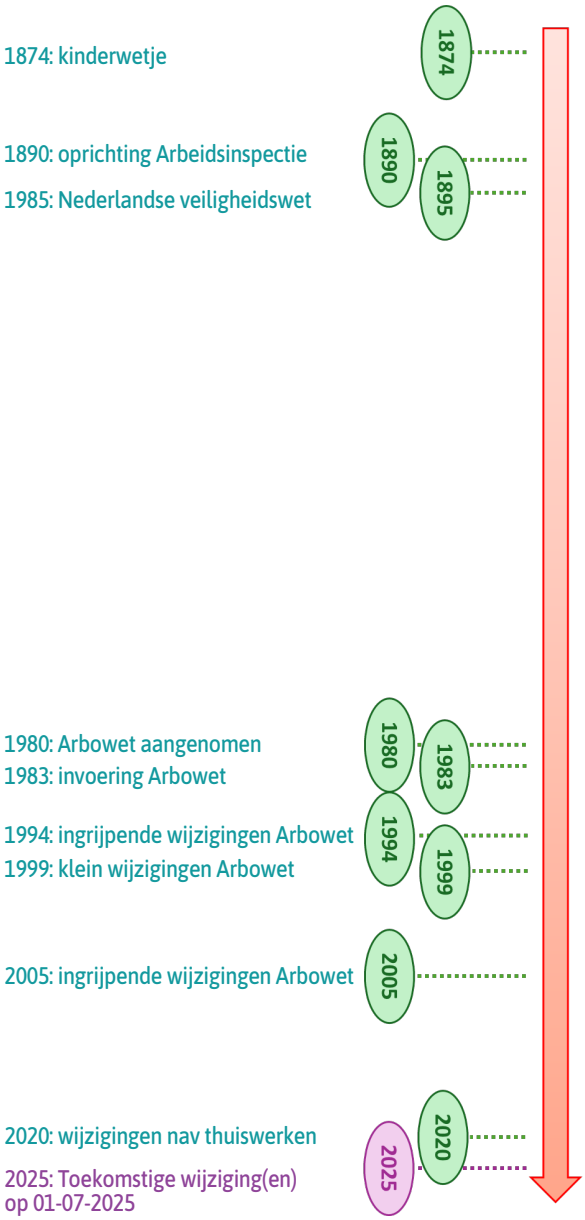
ORGANISATIE

Waarom zijn ergonomen geen arbo-kerndeskundigen?

De Arbeidsomstandighedenwet – of de Arbowet, zoals deze wet tegenwoordig wordt genoemd – onderscheidt vier kerndeskundigen: de bedrijfsarts, de veiligheidskundige, de arbeidshygiënist en de arbeids- en organisatiedeskundige. Ergonomie wordt dus niet in de wet als kerndiscipline erkend en in lijn daarmee is de ergonoom geen kerndeskundige. Gezien de centrale rol van ergonomie in ontwerp en organisatie van werk is dat niet alleen opmerkelijk; het heeft belangrijke gevolgen voor de inbreng van ergonomie in de arbeidsomstandigheden van miljoenen werkenden in Nederland.

Erwin Speklé en Ernst Koningsveld

De arbeidsomstandigheden met betrekking tot ergonomie en fysieke belasting in de Risico-Inventarisatie en -Evaluatie (RI&E) mogen niet worden getoetst door gecertificeerde ergonomen, ondanks het feit dat zij bij uitstek de experts op dit gebied binnen een arbodienst zijn. In plaats daarvan valt dit nu onder de verantwoordelijkheid van arbeidshygiënisten, die weliswaar zeer bekwaam zijn in hun vakgebied, maar over het algemeen aanzienlijk minder kennis hebben van ergonomie en fysieke belasting dan ergonomen. Hoe is dat zo gekomen? Een blik terug in de geschiedenis.



De Arbowet

In 1980 werd de Arbeidsomstandighedenwet door het parlement aangenomen. De Arbowet, zoals die gewoonlijk wordt genoemd, verplicht werkgevers en werknemers om risico's in het werk voor de gezondheid, de veiligheid en het welzijn van werknemers te verminderen. Ook zelfstandig ondernemers zijn verplicht gezondheid en veiligheid van zichzelf en anderen te beschermen. Het doel is om ongevallen en aandoeningen die door het werk worden veroorzaakt, te voorkomen. De wet regelt ook dat werkgevers moeten samenwerken met kerndeskundigen die in de praktijk arbo-deskundigen worden genoemd. Zij adviseren over het verbeteren van de arbeidsomstandigheden en ze controleren de Risico-Inventarisatie en -Evaluatie (RI&E). Om dat werk te mogen doen, moeten kerndeskundigen gecertificeerd zijn. De wet is sinds de inwerkingtreding op meerdere momenten meer of minder ingrijpend aangepast, maar de eerdergenoemde vier kerndisciplines werden steeds gehandhaafd.

Ontstaan van de Arbowet

De eerste wet die iets doet aan het beschermen van werkenden was het Kinderwetje van Van Houten (1874). Deze wet moest vooral een eind maken aan de kinderarbeid. In 1890 is de Arbeidsinspectie opgericht. Niet lang daarna werd de Nederlandse Veiligheidswet van 1895 aangenomen; bedoeld om de veiligheid tijdens het werk te regelen. De wet was erop gericht om werkenden in fabrieken of werkplaatsen te beschermen; in eerste instantie ging het alleen om omgevingen waar krachtwerktuigen of ovens stonden. Via deze wet werd de weg vrijgemaakt voor een toenemend aantal veiligheidsvoorschriften in decreten, besluiten en verordeningen voor bepaalde bedrijfstakken en beroepsgroepen.

In de jaren zeventig van de twintigste eeuw nam het ziekteverzuim flink toe en in lijn daarmee het aantal arbeidsongeschikten. Naast persoonlijk leed leidde dat tot hoge kosten voor bedrijven en de maatschappij. Als maatregel nam de Tweede Kamer in 1980 de Arbeidsomstandighedenwet aan (kort: Arbowet). Het is opmerkelijk dat dit een unaniem besluit

was. Vanaf 1983 wordt de Arbowet stapsgewijs ingevoerd in Nederland. In 1994 is deze wet ingrijpend gewijzigd en in 1999 hebben nog een aantal kleine wijzigingen plaatsgevonden. In 2005 volgde een grootschalige wijziging die onder meer inhoudt dat bedrijven niet langer verplicht zijn zich aan te sluiten bij een Arbodienst. Sindsdien zijn er nog meer aanpassingen doorgevoerd, met als belangrijkste de aanpassing na de coronapandemie in 2020 naar aanleiding van medewerkers die vaker thuis werken.

Hoe kwam het tot vier kerndeskundigen?

In de jaren zeventig werd de Arbowet voorbereid. In dat traject werden meerdere hoorzittingen met deskundigen en belanghebbenden gehouden. Dat de bedrijfsarts en de veiligheidskundige als kerndisciplines zouden worden erkend, werd door niemand betwist. Meerdere andere disciplines lobbyden om ook als kerndiscipline te worden erkend. Het bestuur van de Nederlandse Vereniging voor Ergonomie (NVvE), de voorloper van Human Factors NL, heeft destijds actief gelobbyd om de 'stand van de wetenschap op het gebied van de ergonomie' opgenomen te krijgen in de concepttekst van de Arbowet. Vooral toenmalig NVvE-voorzitter Pieter Rookmaaker zette zich hier met veel energie voor in. Hij vertelt: "Ik herinner me dat de eerste versies van de wet veel meer ontwerpend waren georiënteerd dan het uiteindelijke resultaat. Als NVvE hebben we met politieke fracties in de Tweede Kamer overleg gevoerd. Uiteindelijk is het amendement waarbij ergonomie expliciet bij de kerndisciplines opgenomen zou worden, 'uitgeruild' bij de behandeling van het wetsontwerp voor een andere aanpassing. Ik was aanwezig bij de behandeling in de Tweede Kamer. Er waren meer dan 100 amendementen op de voorgestelde wettekst. Het was 'je reinste koehandel' tussen de verschillende politieke fracties in de geest van: wij gaan akkoord met jullie voorstel als jullie akkoord gaan met ons voorstel!" Op de achtergrond speelde de voor die disciplines 'vanzelfsprekende' opstelling van de bedrijfsartsen en veiligheidskundigen. Zij hadden in de oude wetgeving al een sterke positie, mede dankzij hun sterke beroepsverenigingen: de Nederlandse Vereniging voor Arbeids- en Bedrijfsgeneeskunde en de Nederlandse Vereniging van Veiligheidskundigen. De NVvE was vooral een vereniging van 'in ergonomie geïnteresseerden'. Slechts een deel van de leden had een functie met de naam 'ergonoom'. De NVvE was dus geen beroepsvereniging, wat de positie niet sterk maakte. Bedenk hierbij dat registratie van ergonomen op landelijk, laat staan op Europees niveau, nog toekomstmuziek was. De Stichting Registratie ergonomen (SRe) werd pas in 1991 opgericht en het Center for Registration of European Ergonomists (CREE), samen met de titel European Ergonomist (Eur.Erg.), in 1994.

Andere vakgebieden die naar een positie als kerndeskundige streefden, waren onder meer de bedrijfsverpleegkunde, de arbeids- en organisatiekunde, en de toen nog niet eens zo lang bestaande arbeidshygiëne. Allen bepleitten hun positie, en hadden een mening over de andere disciplines. Rookmaaker: "Het leek wel een slagveld. Daar kwam bij dat de toenmalige directeur-generaal van de Arbeid van mening was dat ergonomie al in de opleiding van veiligheidskundigen en bedrijfsartsen zat, en daarmee afgedekt was. Bovendien wilden sociale partners uit overwegingen van kosten en complexiteit het aantal kerndeskundigen beperkt houden."

Hoe werd dit onder ergonomen beleefd?

Paul Settels, van begin af aan betrokken bij de ontwikkeling en introductie van de registratie van ergonomen, herinnert zich het volgende. "Ergonomen waren en zijn actief in alle sectoren en houden zich daar bezig met sterk verschillende thema's. De ergonomen die binnen de arbodienstverlening werkzaam zijn, hebben naast bescherming ook als doel 'de mens in zijn werk optimaal te laten functioneren'. Uiteraard willen zij, naast de vele andere projecten die ze doen, tenminste helpen voorkomen dat werknemers arbeidsgerelateerde klachten krijgen of zelfs ziek worden door hun werk. Een nobele en belangrijke taak. Ergonomen binnen arbodiensten voorzien zo een aanzienlijk deel van de organisaties en werkende bevolking in Nederland van advies op het gebied van ergonomie en human factors, op alle niveaus.

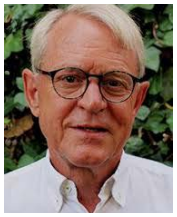
In de jaren zeventig en vroege jaren tachtig waren er zeker ook leden van de NVvE die juist *niet* in de hoek van 'arbo en wetgeving' terecht wilden komen. Zij zagen hun rol vooral om 'de mens in zijn werk optimaal te laten functioneren' en dat vanuit een ontwerpdiscipline; meedenken en mee-ontwerpen, maar beslist niet als 'toezichthouder op niet-ziek-worden van je werk'. We hebben een breed vakgebied dat veel verder gaat dan fysieke en mentale arbeidsbelasting. Die verscheidenheid aan ergonomen hebben we tot op de dag van vandaag."

Hoe nu verder?

Gezien de historische en huidige context van de Arbowet is de rol van de ergonoom hierin nog steeds duidelijk onderbelicht. Dit roept de vraag op of wij als ergonomen deze discussie moeten heropenen. Moeten we niet streven naar een herziening van de wetgeving om de cruciale bijdrage van ergonomie aan een gezonde en veilige werkomgeving te erkennen en te waarborgen? Het is wellicht tijd om onze positie opnieuw te evalueren en te pleiten voor een formele erkenning als kerndeskundige, zodat we onze expertise ten volle kunnen inzetten voor het welzijn van werkenden in Nederland.

Over de auteurs

Erwin Speklé Eur.Erg.
Arbo Unie, Haarlem



Ernst Koningsveld Eur.Erg.
Competentzijn.nl, Oegstgeest

Zien we jullie binnenkort?

Dit jaar biedt onze vereniging volop kansen om elkaar te ontmoeten, (oude) bekenden te zien en nieuwe human factors specialisten te leren kennen. Inmiddels kijken we terug op geslaagde bijeenkomsten over robotisering bij het RoboHouse in Delft en over Zwaar werk in samenwerking met TNO.

Ook de komende tijd staat er nog van alles op de rol:

- 18 juni **Themabijeenkomst** over de relatie van tillen op het werk en rugpijn i.s.m. de VvBN.
- 9 juli **Bedrijfsbezoek** Ergotron in Amersfoort, tevens **Algemene Ledenvergadering**.

Meer over deze bijeenkomsten en mogelijkheid tot inschrijven is te vinden op onze website.

HFNL Jaarcongres

Wat ook te vinden is op onze website is de vernieuwde congrespagina. We sluiten dit volle HFNL-jaar af met ons tweedaagse congres op 27 en 28 november op onze vertrouwde locatie Woudschoten in Zeist.
Het **HFNL-Jaarcongres** heeft dit jaar als congresthema '**Duurzaam**', een onderwerp dat actueler is dan ooit, ook binnen het Human Factors-vakgebied. Gedurende twee dagen onderzoeken we wat duurzaam betekent in relatie tot mens, werk en ontwerp. We brengen professionals, onderzoekers, beleidsmakers en ontwerpers samen rond de vraag: hoe bouwen we aan systemen, organisaties en werkomgevingen die écht toekomstbestendig zijn? En we verkennen het thema duurzaam vanuit verschillende Human Factors-perspectieven.

Duurzame inzetbaarheid **-** Hoe ondersteunen we mensen om gezond, veilig en gemotiveerd aan het werk te blijven?
Duurzaam ontwerp **-** Hoe ontwerpen we producten, technologie en omgevingen die lang meegaan, flexibel zijn én aansluiten bij de gebruiker?
Duurzame zorg **-** Hoe creëren we een zorgsysteem dat mensgericht, efficiënt en houdbaar is?
Duurzame gedragsverandering **-** Hoe stimuleren we gedrag dat bijdraagt aan gezondheid, veiligheid en welzijn – niet alleen op korte termijn, maar blijvend?

Via de congrespagina op humanfactors.nl kunt u zich inschrijven voor het congres, uw interesse in het organiseren van een sessie of spreken op het congres kenbaar maken, of een inzending doen voor een van de HFNL-prijzen. Zoals gezegd, de vernieuwde congreswebsite is inmiddels online. Neem vooral een kijkje.

Wij hopen jullie natuurlijk op één of meerdere van deze bijeenkomsten te treffen.

Het bestuur van Human Factors NL
Marijke Melles, Pieter Coenen en Reinier Hoftijzer