Borgen van eerlijkheid bij het gebruik van AI in de zorg

Wereldwijd zijn er veel initiatieven op het gebied van kunstmatige intelligentie (AI) die gebruik maken van slimme algoritmen, die op basis van trainingsgegevens meer verbanden kunnen leggen dan voor een menselijk brein mogelijk is. Nu er meer AI in gebruik is, rijzen er zorgen over discriminatie van bepaalde individuen of groepen. Het aantal voorbeelden van AI die onbedoeld vooroordelen bevatten of om een andere reden tot oneerlijke uitkomsten leiden, neemt toe.

& DOOR ROB DOMS, GUIDO HUISINTVELD

Een goed voorbeeld: foto-analyserende algoritmen die discriminerend zijn, of die de kans op herhaling van een overtreding bepalen, die mensen met een donkere huidskleur verkeerd classificeren. Voorbeelden van oneerlijke uitkomsten zijn ook te vinden in de geneeskunde, bijvoorbeeld minder nauwkeurigheid bij het identificeren van huidkanker bij personen met een donkere huidskleur als gevolg van ongelijke vertegenwoordiging in trainingsgegevens. Bias in algoritmen is ongewenst, maar regelgeving blijft achter bij de technische ontwikkelingen. Vanwege het ontbreken van wet- en regelgeving zoeken organisaties naar geschikte randvoorwaarden om algoritmen eerlijk te houden. Zij hebben immers de plicht om het risico op discriminatie bij het ontwerp en de toepassing van AI zoveel mogelijk te voorkomen of te verkleinen, zoals ook gevraagd door Amnesty International en Access Now.

Verschillende artikelen beschrijven de noodzaak om te voldoen aan biomedische ethische kaders bij het gebruik van AI in de gezondheidszorg door de vier ethische principes van welwillendheid, niet-schadelijkheid, autonomie en rechtvaardigheid toe te passen. Andere stellen dat deze biomedische ethische kaders niet volstaan en dat ook de oorsprong en uitlegbaarheid van een algoritme gewaarborgd moet zijn.

Toegang tot Wlz

Machine learning-toepassingen zijn ook interessant voor het CIZ. Het CIZ is de organisatie die de toegang bepaalt voor de langdurige zorg (Wlz) en hiermee een cruciale rol speelt tussen wetgeving en uitvoering. Het CIZ bepaalt namelijk, middels de toegang tot de Wlz, de toegang tot financiële middelen voor de zorgverlening die binnen deze wetgeving valt. De instantie behandelt een forse stroom aanvragen per jaar.

Guido Huisintveld, informatiemanager bij het CIZ, heeft in dat kader een masterthesis geschreven als afsluiting van zijn masteropleiding Advanced Health Informatics Practice (AHIP) van hogeschool Inholland. Het lectoraat Medische Technologie heeft Guido begeleid bij het schrijven van publicaties naar aanleiding van deze masterthesis.

Voor verschillende taken van het CIZ is het nodig om na te gaan of alle relevante informatie aanwezig is, wat de verwachte complexiteit is, of er signalen zijn die kunnen wijzen op een frauderisico en welk type onderzoek passend is voor de beoordeling. Dit zijn tijdrovende trajecten waarbij veel data wordt vastgelegd met betrekking tot de uiteindelijke beslissing. Er zijn dus ML-mogelijkheden om ondersteuning te bieden en efficiënter te werken. Voorbeelden zijn de herkenning van documenten en beslissingsondersteuning.

Een ander voordeel van ML is het voorkomen van menselijke vooroordelen. Deze vooroordelen kunnen worden beperkt met beslissingsondersteuning bij het nemen van complexe beslissingen.

Gebrek aan kaders

Door het eerder genoemde gebrek aan kaders en regelgeving zocht Huisintveld naar een ethisch correcte manier om datagedreven besluitvorming over meerdere domeinen toe te passen. Zijn vertrekpunt was het in kaart brengen van de ondersteuningsbehoeften van werknemers.

Dit leidde tot de onderzoeksvraag: welke ondersteuningsbehoeften hebben professionals om te zorgen voor ethische kaders voor het ontwerpen, ontwikkelen en gebruiken van ML, om het risico op vooringenomenheid en oneerlijkheid door ML-algoritmen te minimaliseren? Middels een verkennend literatuuronderzoek om inzicht te krijgen in ethiek, vooroordelen en oneerlijkheid die gepaard kunnen gaan met AI, is een reeks interviewonderwerpen geïdentificeerd. Op basis van de relevante onderwerpen is vervolgens een interviewprotocol opgesteld. De onderzoekspopulatie bestaande uit huidige en voormalige werknemers die betrokken zijn of zijn geweest bij het verkennen van de mogelijkheden voor ML. Dit werd gedaan om een basiskennis van het onderwerp te verzekeren. Via semigestructureerde interviews is de data verzameld.

Tijdens de analyse van de interviews is een aantal belangrijke topics geïdentificeerd en als hoofdcategorie gepositioneerd (zie kader). Deze categorieën bestrijken verschillende belangrijke domeinen die van invloed zijn op het maken en gebruiken van ML en worden gebruikt als een structuur om de resultaten te beschrijven.

Conclusie

Op basis van deze kwalitatieve studie kunnen we concluderen dat de ondersteuningsbehoeften die het risico op vooroordelen en oneerlijkheid helpen verminderen, te vinden zijn op drie gebieden: 'Organisatievisie, beleid en principes', 'ontwerp en ontwikkeling' en 'werken met ML-applicaties'.

Binnen het gebied 'Organisatievisie, beleid en uitgangspunten' hebben medewerkers behoefte aan een duidelijke visie op het gebruik van ML-applicaties en concrete algemene kaders en richtlijnen die bepalend zijn voor de verdere inrichting en het gebruik van de applicaties. Er wordt expliciet op gewezen dat aandacht moet worden besteed aan de transparantie en de borging van menselijke tussenkomst bij de besluitvorming.

In de categorie 'ontwerpen en ontwikkelen' is er behoefte aan ondersteuning door externe expertise zodra bevindingen uit datalabs leiden tot de wens tot verdere ontwikkeling en implementatie. De geraadpleegde expert(s) kunnen helpen voorzien in de ondersteuningsbehoefte door middel van technische kaders en richtlijnen, mogelijkheden om het algoritme te monitoren en mogelijkheden om het algoritme bij te werken.

In het domein 'werken met ML-applicaties' is behoefte aan verdere ondersteuning met kennis over de werking van het algoritme, het kunnen begrijpen van de resultaten en het gebruik van deze uitkomsten van ML-applicaties. Daarnaast is er behoefte aan een toename van het databewustzijn om de datakwaliteit te verbeteren. ■

Topic	Toelichting
Functie en ervaring met het onderwerp.	De functie kan mogelijk interessant zijn in relatie tot de aanwezige kennis over het onderwerp en daarmee de behoefte in ondersteuning. Ook de ervaring die de geïnterviewde heeft kan hierop van invloed zijn.
Bekendheid met problematiek van AI/ machine learning.	Het is goed om te weten of de medewerker zich bewust is van de risico's van machine learning en artificial intelligence. Aangezien dit van invloed kan zijn op de ondersteuningsbehoefte bij het eerlijk inrichten van de machine learning.
Voorbereiding van het project.	Het voorkomen van vooroordelen in het algoritme begint met de ontwerpfase. Inzichten in ethische risico's bij het ontwerp zijn dan ook relevant. Bijvoorbeeld door vooraf te bepalen welke mate van 'eerlijkheid' gerealiseerd dient te worden of door zorg te dragen voor een multistakeholder aanpak.
De dataset voor het trainen.	De dataset is een belangrijk startpunt van het algoritme. Veel literatuur die gaat over vooroordelen in AI beschrijft de relevantie van de dataset. Zoals de kwaliteit van de data: 'garbage in is garbage out'. Een ander belangrijk punt is dat ook voor alle minderheden voldoende gegevens aanwezig moeten zijn om 'Minority bias' te voorkomen.
De dataset voor het testen.	Het vormgeven van testdata is ook een belangrijk aspect om te komen tot eerlijke algoritmes. Om deze testdata te laten testen of een algoritme eerlijk werkt, is de samenstelling van belang. Als de testdataset een klein geselecteerd gedeelte van de trainingsdata is, zal deze dezelfde vooroordelen bevatten.
Werken met het algoritme.	De wijze waarop gewerkt wordt met het algoritme is een relevant onderdeel voor de eerlijkheid. In de literatuur wordt bijvoorbeeld beschreven hoe het niet afwijken van een algoritme de fouten in het algoritme in stand houdt.
Transparantie en uitlegbaarheid.	Transparantie wordt gezien als belangrijk kader voor eerlijke AI. Hierbij gaat het zowel om transparantie binnen de organisatie als naar burgers toe. Het ontbreken van uitlegbaarheid vergroot het risico op oneerlijke ML.
Visie medewerker over eerlijkheid van huidige algoritme.	Dit raakt de kern van het onderzoek. Hoe vindt de medewerker dat de aandacht voor eerlijkheid van het algoritme nu is en op welke wijze zou deze versterkt kunnen worden?

Guido Huisintveld is een ervaren informatiemanager met een zorgachtergrond, die werkzaam is bij het CIZ.

Rob Doms is teamleider en docent van de masteropleiding Advanced Health Informatics Practice bij Hogeschool Inholland.