

Vertrouwen in de geneeskunde en kunstmatige intelligentie

Lily Eva Frank en Michal Klincewicz

Kunstmatige intelligentie (AI) en systemen die met *machine learning* (ML) werken, kunnen veel onderdelen van het medische besluitvormingsproces ondersteunen of vervangen. Ook zouden ze artsen kunnen helpen bij het omgaan met klinische, morele dilemma's. AI/ML-beslissingen kunnen zo in de plaats komen van professionele beslissingen. We betogen dat dit belangrijke gevolgen heeft voor de relatie tussen een patiënt en de medische professie als instelling, en dat dit onvermijdelijk zal leiden tot uitholling van het institutionele vertrouwen in de geneeskunde.

Machine learning en kunstmatige intelligentie in de geneeskunde

AI/ML-gestuurde systemen kunnen veel impact hebben op het medische besluitvormingsproces zoals diagnoses, prognoses, of de keuze van een behandeling. Daarnaast kunnen deze systemen, zoals aangegeven door Gabriels (2021), artsen helpen met het omgaan met morele dilemma's in de klinische praktijk, bijvoorbeeld als de plicht om de persoonlijke informatie van de patiënt te respecteren in strijd is met de plicht om te handelen in het belang van de patiënt. In dergelijke situaties kunnen AI/ML worden opgevat als ethische ondersteuning in de vorm van een kunstmatige morele adviseur (Giubilini & Savulescu, 2018; Klincewicz, 2016).

Op al deze gebieden (en mogelijk andere) kunnen zij medische professionals vervangen die hun opleiding, ervaring en professionele intuïtie gebruiken om beslissingen te nemen maar ook de gezamenlijke beslissingen die genomen worden met de patiënten en/of andere leden van het medische personeel. Hieronder vallen ook juridische adviseurs en mensen die zitting hebben in speciale besluitvormingsorganen, zoals bijvoorbeeld ethische commissies. Met andere woorden, AI/ML-beslissingen kunnen mogelijk in de plaats komen van professionele beslissingen.

Dit heeft twee belangrijke consequenties voor de relatie tussen een patiënt en de medische wereld als instelling die samenhangen met de uitlegbaarheid en mogelijke fouten van AI/ML-gestuurde besluitvorming. Dit zal onvermijdelijk leiden tot een afbraak van het institutionele vertrouwen in de geneeskunde.

Vertrouwen van patiënten in individuele medische professionals

Patiënten hebben vaak niet de opleiding, ervaring, of intuïtie om medisch advies en medische beslissingen te begrijpen. Zelfs wanneer klinische informatie goed wordt gecommuniceerd en begrepen, worden deze beslissingen beïnvloed door de juridische, ethische en culturele context van de arts of van het ziekenhuis. Deze asymmetrie van kennis betekent voor de zorgverlening dat patiënten erop zullen moeten vertrouwen dat de medische professional handelt in hun belang en dat de professional de normen en waarden van de patiënt respecteert. Dit schept een speciale set van plichten voor medische professionals, waarvan de belangrijkste zijn dat ze als professionals betrouwbaar zijn en het ook inderdaad verdienen om dat vertrouwen te krijgen (Rhodes, 2020). Dat is het geval als patiënten hen dat vertrouwen omtrent medische beslissingen geven en ook laten zien. Natuurlijk zijn er ook omstandigheden waarin de arts te vertrouwen is maar de patiënt hem toch niet vertrouwt.

Asymmetrie van kennis betekent voor de zorgverlening dat patiënten erop zullen moeten vertrouwen dat de medische professional handelt in hun belang

Het vertrouwen van de patiënt in het tijdperk van kunstmatige intelligentie

Het vertrouwen van de individuele patiënt in een medische professional of beslissing wordt beïnvloed door het maatschappelijke vertrouwen in de medische wereld als geheel. Dit meer algemene vertrouwen is het resultaat van de mate waarin medische professionals, als groep, in het verleden aan hun plichten hebben voldaan (Rhodes, 2020). In dat geval zullen toekomstige patiënten eerder vertrouwen hebben, zo niet, dan zal het vertrouwen lager zijn (Jacobs et al., 2006).

Dit algemene niveau van institutioneel vertrouwen in de geneeskunde kan echter worden ondermijnd door het *onkritische* gebruik van AI/ML in de geneeskunde. Ten eerste zal in een klinische setting een medische professional de prestaties van een AI/ML-gestuurde medische aanbeveling of beslissing moe-

ten kunnen uitleggen en deze ook kunnen rechtvaardigen. Als bijvoorbeeld een diagnose van kanker gedeeltelijk door een algoritme is vastgesteld, dan zal de patiënt vermoedelijk willen weten waar die diagnose op gebaseerd is en kan hij of zij ook vragen hoe de AI/ML werkt en waarom de arts de diagnose vertrouwt. De arts heeft daarbij de plicht om eerlijk te zijn over de rol van de AI/ML en zijn of haar eigen beperkte begrip van de werking ervan.

Een deel van dit probleem vloeit voort uit de ondoorzichtige of *black-box* aard van de algoritmes in *machine learning* systemen die adviseren over diagnose, prognose en behandelingen.

Inzichten die een AI/ML verschaft zijn niet gemakkelijk in begrijpelijke taal uit te leggen aan de doorsnee patiënt. Dit in tegenstelling tot andere diagnose-instrumenten, zoals CT-scans en bloedtesten, die een interpretatie door een medische professional zelf vergen, wat ook onderdeel is van het klinische besluitvormingsproces. In het geval van een AI/ML-besluit wordt de patiënt gevraagd erop te vertrouwen dat het vertrouwen van de medische professional in het AI/ML besluit gerechtvaardigd is. Alhoewel vertrouwen niet zomaar inwisselbaar is, moeten we ons afvragen of deze overdracht van het vertrouwen van de patiënt in de medische professional naar het AI/ML-systeem niet ten koste kan gaan van het institutionele vertrouwen in de medische professie zelf. Een verkoper die ons merkproducten van hoge kwaliteit verkoopt, zal niet noodzakelijk ons vertrouwen in verkopers in het algemeen verhogen, maar wel in het merk en het product. Op vergelijkbare wijze zal het overdragen van de besluitvorming naar AI/ML het algemene vertrouwen in deze systemen en de ingenieurs die ze ontwerpen, vergroten, ten koste van het vertrouwen van de maatschappij in het medische vakgebied.

De arts heeft de plicht om eerlijk te zijn over de rol van de AI/ML en zijn of haar eigen beperkte begrip van de werking

Vertrouwen en fouten door kunstmatige intelligentie

Een ander gevolg van de introductie van kunstmatige intelligentie met betrekking tot het vertrouwen in de geneeskunde, is dat AI/ML-gestuurde besluitvorming, ondanks de belofte dat zij wellicht nauwkeuriger is dan uitsluitend menselijke besluitvorming, kan leiden tot fouten met schadelijke gevolgen voor de patiënt, zoals een misdiagnose, een fout recept, of gewoon een gebrek aan gevoeligheid voor de waarden van de patiënt. Fouten kunnen tot juridische consequen-

ties leiden zoals rechtszaken en financiële sancties. Maar hoe zal de juridische (en morele) verantwoordelijkheid worden vastgesteld in gevallen waar AI/ML-gestuurde besluitvorming bij betrokken was? Zal de medische professional verantwoordelijk worden gehouden? Of het bedrijf dat de AI/ML-systemen leverde? Dit is een specifieke versie van een algemeen probleem in ethiek van robotica en kunstmatige intelligentie: de verantwoordelijkheidskloof (Matthias, 2004; de Sio & Mecacci, 2021). Dit is het fenomeen waarbij een onkritische invoering van automatisering in complexe besluitvormingsprocessen het onmogelijk maakt om verantwoordelijkheid te leggen bij een individu. Een belangrijk gevolg hiervan is de afbraak van vertrouwen omdat van AI/ML-ingenieurs, programmeurs en bedrijven over het algemeen niet wordt verwacht dat zij hetzelfde niveau of dezelfde soort morele verplichtingen hebben als medische beroepsbeoefenaren.

Behoud van het vertrouwen in de geneeskunde als beroep

Algemeen wordt ervan uitgegaan dat het belang van de patiënt voorop staat bij de medische professional en dat dit de medische institutie kenmerkt. Maar als AI/ML erbij betrokken wordt, dan kan de medische professional geen verantwoordelijkheid nemen over fouten en beslissingen, omdat de medische professional geen ingenieur is en uiteindelijk niet verantwoordelijk is voor het juist functioneren van het AI/ML-systeem. Door deze complexe situatie is het lastig om met voldoende zekerheid te generaliseren over toekomstige juridische procedures of ethische oordelen. Er moet daarom worden nagedacht over de mate waarin de arts misschien gedeeltelijk verantwoordelijk kan worden gesteld hiervoor.

Er moet worden nagedacht over de mate waarin de arts misschien gedeeltelijk verantwoordelijk kan worden gesteld

Er is echter een groot verschil tussen de toewijzing van verantwoordelijkheid voor misdiagnoses bij gebruik van AI/ML en situaties waar alleen professionals bij betrokken zijn. In dat laatste geval wordt gebruik gemaakt van het institutionele vertrouwen van mensen in de beroepsgroep om medische professionals niet aansprakelijk te stellen (moreel of juridisch) voor ingrepen waar dat in andere contexten wel het geval zou zijn geweest, zoals snijden, “vergiftigen” (medicijnen toedienen), of verkeerd informeren. In het geval van AI/ML beslissingen ligt de aansprakelijkheid in eerste instantie bij de professional vanwege zijn vertrouwen in het AI/ML-systeem, maar dat roept het beeld op dat de professio-

nal zelf niet betrouwbaar is. Als er meer van dit soort situaties zich voordoen, kan het vertrouwen in de medische beroepsbeoefenaars als groep worden aangetast.

Om de onvermijdelijke erosie van institutioneel vertrouwen in de geneeskunde tegen te gaan moeten we sterk benadrukken dat de beoefening van geneeskunde meer is dan het sleutelen aan het lichaam om de optimale werking te behouden, alsof het een auto is. Het beoefenen van geneeskunde gaat gepaard met rechten en plichten voor professionals, die anderen niet hebben. Zoals Rosamond Rhodes opmerkt kunnen artsen in een lichaam van iemand snijden, medicatie verstrekken, en onze naakte lichamen onderzoeken, maar tegelijkertijd moeten zij zich ook aan de geheimhoudingsplicht houden, handelen in het belang van anderen, eerlijk zijn, en streven naar vertrouwen en dat vertrouwen ook verdienen (2020). Dit alles op een manier die veel verder gaat dan wat we verwachten van mensen in andere beroepen. De geneeskunde is een fundamenteel moreel toegewijd beroep. De introductie van door AI/ML ondersteunde besluitvorming in de klinische situatie moet zorgvuldig beoordeeld worden in het licht van deze bijzondere plichten en rechten.

We moeten een manier vinden waarop we het institutionele vertrouwen, dat in het hart van de medische praktijken ligt, kunnen beschermen, aangezien dat vertrouwen het mogelijk maakt voor patiënten om zichzelf kwetsbaar op te stellen tegenover interventies die zij misschien niet volledig kunnen waarderen of begrijpen. Als we dit niet doen, kunnen medische professionals en de geneeskunde als instelling hun speciale status, en het hoge vertrouwen dat ze krijgen, verliezen. Om AI/ML op een ethische manier in de medische praktijk te integreren, is op zijn minst een aanvullende medische (en voortgezette) opleiding nodig. Dit leerplan moet worden ontwikkeld in samenwerking met artsen, patiëntvertegenwoordigers, ethici en AI/ML-wetenschappers.

Dr. Lily Eva Frank is assistant professor Industrial Engineering and Innovation Science bij de Philosophy and Ethics Group van Eindhoven University of Technology.

Dr. Michal Klincewicz is assistant professor Cognitive Science and Artificial Intelligence bij Tilburg University.

Literatuur

-
- De Sio, F.S. & Mecacci, G. (2021). Four responsibility gaps with artificial intelligence: Why they matter and how to address them. *Philosophy & Technology*, <https://doi.org/10.1007/s13347-021-00450-x>
- Giubilini, A. & Savulescu, J. (2018). The artificial moral advisor. The “ideal observer” meets

artificial intelligence. *Philosophy & Technology*, 31(2), pp. 169-188.

Jacobs, E.A., Rolle, I., Ferrans, C.E., Whitaker, E.E. & Warnecke, R.B. (2006). Understanding African Americans' views of the trustworthiness of physicians. *Journal of General Internal Medicine*, 21(6), pp. 642-647.

Klincewicz, M. (2016). Artificial intelligence as a means to moral enhancement. *Studies in Logic, Grammar and Rhetoric*, 48(1), pp. 171-187. <https://doi.org/10.1515/slgr-2016-0061>

Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology*, 6(3), pp. 175-183. <https://doi.org/10.1007/s10676-004-3422-1>.

Rhodes, R. (2020). *The trusted doctor: Medical ethics and professionalism*. Oxford: Oxford University Press.