

## AI WILL SEE YOU NOW

Het perspectief van de arts en de patiënt op een verantwoorde toepassing van Artificiële Intelligentie (AI) ter ondersteuning van medische beslissingen

Dr. Joanna E. Klopotowska

**SIDN**fonds

## **PROJECTTEAM**

**Prof. dr. Ameen Abu-Hanna** – hoogleraar medische informatiekunde, Amsterdam UMC

**Prof. dr. Ellen Smets** – hoogleraar medische psychologie, Amsterdam UMC

**Prof. dr. Nicolette de Keizer** - hoogleraar medische informatiekunde, Amsterdam UMC

**Prof. dr. Dick Willems** – hoogleraar medische ethiek, Amsterdam UMC

**Dr. Dave Dongelmans** – intensive care arts en senior onderzoeker, Amsterdam UMC

**Dr. Danielle Sent** – senior docent en onderzoeker klinische informatiekunde, Amsterdam UMC

**Dr. Joanna Klopowska** – senior onderzoeker klinische informatiekunde en ziekenhuisapotheker (projectleider), Amsterdam UMC

**Irene Stellingwerf, MSc** – onderzoeksassistent, Amsterdam UMC

## DEELNEMERS



UMC Utrecht



umcg

ZIEKENHUIS

ST ANTONIUS



UMC Utrecht  
Julius Centrum



pacmed



waag society

institute for art, science and technology

*“Technology can only find its place in human life  
if it meets the actual needs of the user.”*

Georg Wilhelm Friedrich Hegel 1770-1831

## INHOUDSOPGAVE

Projectteam	Pagina 2
Deelnemers	Pagina 3
Voorwoord	Pagina 6
Afbakening AI in het project	Pagina 7
Samenvatting	Pagina 8
Inleiding en Methode	Pagina 10
Resultaten	Pagina 11
Slotwoord	Pagina 34
Referenties	Pagina 35

## VOORWOORD

**Artificiële intelligentie (afgekort AI)**, kent haar oorsprong als wetenschappelijke discipline vanaf 1956, en heeft verschillende fases van optimisme maar ook teleurstellingen ervaren. In het laatste decennium spreekt men van een doorbraak in een subdiscipline van AI die Machine Learning heet. Dit komt vooral door sterk toegenomen rekenkracht van de computers en grote toename in beschikbaarheid van digitale gegevens. Iedereen heeft tegenwoordig op kleine of grote schaal met AI te maken, omdat de toepassing hiervan via smartphone en private en publieke diensten haar intrede heeft gemaakt. Echter, deze doorbraak gaat ook gepaard met AI-hypevorming. Vooral de sociale media staan vol van berichten over wonderlijke AI-toepassingen zonder enkele of met slechts zeer gebrekkige wetenschappelijke onderbouwing. Ook zijn er inmiddels meer dan 84 ethische richtlijnen en gedragscodes over AI gepubliceerd, omdat men beseft dat een AI-toepassing verantwoord moet zijn. Wat dat betekent blijft echter vaag en niet actiegericht. Concretisering is noodzakelijk maar hiervoor dient geleerd te worden van specifieke case-studies in concrete domeinen.

In dit rapport brengen we **concrete behoeften en wensen van zorgverleners, patiënten en technologie experts** in kaart ten aanzien van toepassing van AI-technologie ter ondersteuning van medische beslissingen. De verkregen inzichten zijn vervolgens vertaald naar randvoorwaarden voor **verantwoorde toepassing van AI** vanuit het perspectief van de belangrijkste stakeholders. Deze **randvoorwaarden** kunnen gebruikt worden tijdens verschillende fasen in projecten waarin beslissingsondersteuning door AI-technologie wordt beoogd. Dit rapport is interessant voor patiënten, zorgverleners, beleidsmakers en managers in de gezondheidszorg, zowel op nationaal als organisatieniveau.

AI-technologie heeft de potentie om de manier waarop we ziekten voorkomen, diagnosticeren en behandelen drastisch te transformeren. Echter, **AI-technologie is geen magic bullet**. Wat we niet moeten vergeten is dat de gegevens die in zorgsystemen opgeslagen zitten en verzameld worden, gegevens zijn die ten eerste niet verzameld zijn of worden met als doel het ontwikkelen van AI-modellen. Daarnaast zijn deze gegevens grotendeels handmatig ingevoerd door mensen of door tussenkomst van mensen ontstaan. Hierdoor kunnen diverse datakwaliteitsbeperkingen en bias in deze gegevens aanwezig zijn. Het zijn ook mensen die beslissingen nemen over hoe deze gegevens voor te bereiden voor data-analyses, hoe de computer met deze gegevens om moet gaan en hoe de ontwikkelde AI-modellen in de praktijk worden getest, geëvalueerd en uiteindelijk gebruikt. Deze beslissingen hebben een sterke invloed op de betrouwbaarheid, klinische meerwaarde, schaalbaarheid en generaliseerbaarheid van AI-modellen. Het is daarom niet alleen de AI-technologie aan zich, maar **de interactie tussen AI-technologie en de mens**, die mede bepalend zal zijn of de toepassing van AI-technologie ter ondersteuning van medische beslissingen ten goede uitpakt voor onze maatschappij.

Onze dank gaat uit naar alle deelnemers en alle betrokken adviserende experts voor hun medewerking aan dit onderzoek en het SIDN-fonds voor hun financiële steun aan dit onderzoek. In het bijzonder willen we Marike van der Schaaf – associate professor Acute Care Rehabilitation - bedanken voor haar hulp bij het werven van IC-patiëntvertegenwoordigers voor dit onderzoek.

### **Prof. dr. Ameen Abu-Hanna**

Hoogleraar Medische Informatiekunde met expertise in artificiële intelligentie, prognostische modellen en geautomatiseerde beslissingsondersteuning  
AMC-UvA

Principal Investigator  
Klinische Informatiekunde, Amsterdam UMC

### **Prof. dr. Nicolette de Keizer**

Hoogleraar Medische Informatiekunde in het bijzonder evaluatie van zorg en zorginformatiesystemen  
AMC-UvA

Afdelingshoofd en Principal Investigator  
Klinische Informatiekunde, Amsterdam UMC

## AFBAKENING VAN AI EN HAAR TOEPASSING IN DIT PROJECT

In dit project beperken we ons tot **supervised machine learning** als we het hebben over AI. Supervised machine learning is op dit moment een van de meest toegepaste vormen van AI en is gericht op het genereren van kennis uit data door de computer. Hierbij worden statistische methoden gebruikt om bijvoorbeeld een voorspellend model te leren. Het maken van een model door een computer kan gezien worden als het vertalen van een bepaald probleem in wiskundige termen. Zo'n model kan een uitkomst voorspellen, zoals sterfte, aan de hand van ontdekte verbanden in de gegevens. Dit doet de computer aan de hand van voorbeelden die de mensen hem aanbieden (zoals patiëntgegevens over aandoeningen, klachten, symptomen, medicatie, uitslagen van onderzoeken). Het belangrijkste verschil tussen machine learning en standaard statistische voorspellende modellen is dat de vorm (ook wel formule) van het geleerde voorspellende model door machine learning niet van tevoren vast staat, zoals het geval is bij de standaard statistische modellen. Dit betekent dat machine learning vrij flexibele voorspellende modellen kan leren die betere voorspellingen geven dan de wat starre standaard statistische modellen. Dit is met name het geval wanneer er veel gegevens in een domein aanwezig zijn om van te leren.

Verder hebben we als gespreksonderwerp in dit project gekozen voor een **voorspellend AI-model die de kans op overleving van Intensive Care opname** berekent. Op de afdeling Intensive Care (afgekort IC) worden patiënten behandeld met vaak levensbedreigende aandoeningen. Zorg op een IC-afdeling behoort tot de duurste zorg in het hele ziekenhuis, waarbij kosten van IC-zorg geschat worden op ongeveer 20% van het totale ziekenhuisbudget. De totale kosten per IC-dag per patiënt bedragen circa € 2500. De emotionele kosten van een IC-behandeling zijn ook erg hoog, waarbij patiënten en hun naasten niet noodzakelijkerwijs de voorkeur geven aan intensieve, levensbehoudende IC-behandeling. Regelmatig worden op een IC-afdeling patiënten opgenomen met een lage overlevingskans en/of lage verwachte kwaliteit van leven na een IC-opname. Een behandeling is als medisch zinloos te beschouwen als herstel naar een zinvol bestaansniveau niet meer is te verwachten. Zinloos doorbehandelen is niet respectvol tegenover de patiënt, diens naasten en het behandelteam. Daarom is juist voor de IC-afdeling de vraag relevant of het uitvoeren van alle complexe behandelingen altijd wel even zinvol is. Niet alleen vanwege de kosten, maar juist ook in het belang van de patiënt. Een betrouwbare voorspelling kunnen maken van de kans op overleving met acceptabel geachte kwaliteit van leven na een IC-opname en in overeenstemming met de voorkeuren van de patiënt of de naasten, is dus van essentieel belang voor optimale patiëntenzorg.

## SAMENVATTING

In dit rapport gaan we in op wat verantwoord gebruik van AI-technologie betekent vanuit het perspectief van de arts en patiënt. Het toepassen van AI-technologie ter ondersteuning van medische beslissingen heeft sinds circa 10 jaar grote vlucht genomen. Supervised machine learning is op dit moment één van de meest toegepaste vormen van AI. Hierbij worden grote hoeveelheden van patiëntgegevens geanalyseerd en worden verbanden gezocht tussen een uitkomst (bijvoorbeeld sterfte) en input (bijvoorbeeld patiëntkarakteristieken, behandelingen en laboratoriumuitslagen). De output is een wiskundig model die de gevonden verbanden beschrijft en waarmee een voorspelling gedaan kan worden over kans op een uitkomst. Dergelijke voorspellingen worden nuttig geacht om artsen te ondersteunen bij het nemen van medische beslissingen. Omdat de impact van voorspellende AI-modellen op de behandeling van patiënten aanzienlijk kan zijn, is grote zorgvuldigheid getracht bij de ontwikkeling, evaluatie en implementatie van voorspellende AI-modellen in de klinische praktijk. Indien dat niet of onvoldoende gebeurt, kunnen waarden in de gezondheidszorg als toegankelijkheid, betaalbaarheid, diversiteit, autonomie, kwaliteit en vertrouwen, op het spel komen te staan.

Het doel van dit project was daarom om zorgen, behoeften en wensen van zorgverleners en patiënten ten aanzien van toepassing van AI ter ondersteuning van medische beslissingen in kaart te brengen, om vervolgens de verkregen inzichten te vertalen naar randvoorwaarden voor verantwoorde toepassing van AI.

Als gespreksonderwerp in dit project hebben we gekozen voor een voorspellend AI-model dat de kans op overleving van Intensive Care opname berekent. Op de afdeling Intensive Care (afgekort IC) worden regelmatig patiënten opgenomen met een lage overlevingskans en/of lage verwachte kwaliteit van leven na een IC-opname. Daarom is juist voor de IC-afdeling de vraag relevant of het uitvoeren van alle complexe, voor de patiënt belastende, en dure behandelingen zinvol is. Uit eerder onderzoek blijkt dat het nemen van een dergelijke beslissing door het behandelteam echter met veel onzekerheid gepaard gaat. Door inzet van AI kunnen, in vergelijking met mensen, grote hoeveelheden aan gegevens betrouwbaarder, sneller en meer gestandaardiseerd worden geanalyseerd. Op die manier zou een voorspellend AI-model mogelijk van meerwaarde waarde kunnen zijn voor IC-artsen, IC-patiënten en/of naasten van IC-patiënten.

Voor het doel van het project zijn in totaal 10 semigestructureerde interviews uitgevoerd met IC-artsen, 5 met IC-patiëntvertegenwoordigers en 2 met technologie experts. Aansluitend werd een focusgroep bijeenkomst georganiseerd met een afvaardiging van deelnemers aan deze interviews, het projectteam en een aantal externe experts. Voorafgaand aan de interviews zijn 5 experts geraadpleegd en is er een systematische literatuurstudie uitgevoerd om interviewvragen vast te stellen. Op basis van de interviews en focusgroep bijeenkomst, is een set van 22 randvoorwaarden vastgesteld die de zorgen, behoeften en wensen van het merendeel van deelnemers weerspiegelen. Deze randvoorwaarden konden worden ingedeeld in 6 thematische groepen:

1. Randvoorwaarden ten aanzien van **welke medische beslissingen** te ondersteunen met voorspellende AI-modellen.
2. Randvoorwaarden ten aanzien van de **kwaliteit van (medische) gegevens** waarmee voorspellende AI-modellen gemaakt worden.
3. Randvoorwaarden ten aanzien van **ontwikkelaars** van voorspellende AI-modellen.
4. Randvoorwaarden ten aanzien van het **ontwikkeling en evaluatie** van voorspellende AI-modellen.
5. Randvoorwaarden ten aanzien van **gebruik** voorspellende AI-modellen in de klinische praktijk.
6. Randvoorwaarden ten aanzien van **communicatie** van voorspellingen van AI-modellen aan de patiënten.

Hieronder volgt een korte toelichting per thema.



**Randvoorwaarden ten aanzien van welke medische beslissingen te ondersteunen met voorspellende AI-modellen.**

*Voorafgaand aan de ontwikkeling van een voorspellend AI-model dient een gedegen context (business) analyse uitgevoerd te worden met alle betrokken stakeholders. Het doel van zo'n analyse is het in kaart brengen van alle facetten van een medische beslissing, om vervolgens na te gaan of het beoogde voorspellende AI-model daadwerkelijk in de behoeften en wensen van zorgverleners en patiënten (en/of hun naasten) zou kunnen voorzien.*

**Randvoorwaarden ten aanzien van (medische) gegevens die gebruikt worden in voorspellende AI-modellen.**

*Het is van cruciaal belang dat de gegevens die gebruikt worden door voorspellende AI-modellen van voldoende kwaliteit zijn (garbage in, is garbage out). Tevens dienen deze gegevens in voldoende mate de diversiteit van patiënten en settings te weerspiegelen, moeten ze meetbaar zijn en tijdig beschikbaar. Omdat bij toepassing van AI grote hoeveelheden aan diverse patiëntgegevens nodig zijn, is het borgen van privacy van patiënten een andere belangrijke randvoorwaarde. Het checken van vertekening in de gegevens moet ervoor zorgen dat ongewenste patronen worden gepropageerd door AI-modellen.*

**Randvoorwaarden ten aanzien van ontwikkelaars van voorspellende AI-modellen.**

*Het ontwikkelen van voorspellende AI-modellen om medische beslissingen te ondersteunen moet een resultaat zijn van een multidisciplinair teamwork. Participatie van zorgverzekeraars en commerciële partijen wordt als risicovol beschouwd omdat doelstellingen en fiduciaire verplichtingen van deze partijen niet zondermeer op één lijn liggen met die van zorgverleners en patiënten. De ontwikkelaars dienen transparant te zijn over beperkingen in de gebruikte data en daaruit voortvloeiende mogelijke bias in voorspellingen. Eenduidige en eerlijke afspraken tussen alle betrokkenen ten aanzien van eindverantwoordelijkheid, vergoedingen, intellectueel eigendom, moeten voorkomen dat de kwaliteit en toegankelijkheid van de ontwikkelde AI-modellen in het geding komt.*

**Randvoorwaarden ten aanzien van ontwikkeling en evaluatie van voorspellende AI-modellen.**

*Zorgverleners en patiënten eisen gedegen wetenschappelijke bewijsvoering voor effectiviteit van voorspellende AI-modellen. Aan de hand van diverse richtlijnen kan de wetenschappelijke kwaliteit van voorspellende modellen (inclusief AI-modellen) kritisch worden beoordeeld. Om van elkaar te kunnen leren en om vertekeningen in voorspellingen tegen te gaan, worden voorspellende AI-modellen bij voorkeur multicenter ontwikkeld en getoetst. Om de principes van Evidence Based Medicine (EBM) in stand te kunnen houden, moet de zorgverlener in staat worden gesteld de voorspellingen van AI-modellen te kunnen interpreteren en te kunnen beoordelen in hoeverre het van toepassing is op een situatie van een individuele patiënt. Om vervolgens samen met de patiënt tot een gedeelde besluitvorming te komen. De ontwikkelde voorspellende AI-modellen worden als open source modellen beschikbaar gesteld.*

**Randvoorwaarden ten aanzien van gebruik voorspellende AI-modellen in de klinische praktijk.**

*De uitkomsten van voorspellende AI-modellen dienen enkel gebruikt te worden als hulpmiddel om medische beslissingen te ondersteunen. Een menselijke eindbeslissing moet altijd leidend blijven. Door voorspellende AI-modellen in te bouwen in het EPD, zijn de uitkomsten op een toegankelijk en up-to-date wijze beschikbaar voor zorgverleners en patiënten. Eindgebruikers dienen over voldoende kennis en vaardigheden te beschikken om op een verantwoorde manier voorspellende AI-modellen toe te passen. Na implementatie van voorspellende AI-modellen in de klinische praktijk, dient men continu op zowel meerwaarde als (potentiële) risico's te monitoren.*

**Randvoorwaarden ten aanzien van communicatie van voorspellingen van AI-modellen aan de patiënt.**

*De presentatie van voorspellingen van AI-modellen moet aansluiten op het niveau van begrip, de gezondheidsvaardigheden en individuele wensen van patiënten en/of hun naasten. Ziekenhuizen moeten transparant zijn over het gebruik van voorspellende AI-modellen ter ondersteuning van medische beslissingen.*

## INLEIDING

Op dit moment worden beslissingen over wel of niet doorgaan met levensbehoudende IC-behandeling vooral gebaseerd op impliciete kansen, zoals geschat door een multidisciplinair IC-team. Uit onderzoek van Soliman et al. blijkt dat het nemen van een dergelijke beslissing door het behandelteam met veel onzekerheid gepaard gaat. Hoewel artsen momenteel redelijk accurate voorspellingen maken van de prognose van IC-patiënten, zijn hun inzichten beïnvloedbaar door verschillen in ervaring, expertisegebied en hun persoonlijke normen en waarden wat betreft de zorg rondom levensverlengende of beëindigende handelingen. Hierdoor schatten artsen de **kans op overleving van IC-opname** regelmatig te hoog in en soms te laag. In onderzoek van Reader et al. kwam naar voren dat IC-artsen vaak hun besluiten baseren op grond van herinneringen van vergelijkbare situaties uit het (recente) verleden en minder door een analyse van alle beschikbare gegevens en afweging van verschillende opties en kansen. Hierdoor ontstaat veel variatie in de besluitvorming, omdat elke IC-arts andere herinneringen heeft en deze anders weegt. De beïnvloeding van beslissingen door herinneringen nam toe naarmate IC-ervaring langer was.

Door inzet van AI kunnen, in vergelijking met mensen, grote hoeveelheden aan gegevens betrouwbaarder en sneller worden geanalyseerd. Bovendien zou het gebruik van een voorspellend AI-model een dergelijke beslissing meer kunnen standaardiseren, omdat de berekeningen van een computer minder beïnvloedbaar zijn door bovengenoemde menselijke factoren (waardoor minder variatie ontstaat). Echter, om voorspellende AI-modellen op een veilige, doelmatige en kwalitatief hoogwaardige wijze te ontwikkelen en implementeren, waarbij gezamenlijke waarden en normen van zorgverleners en patiënten gerespecteerd worden, is inzicht in hun behoeften, zorgen en wensen ten aanzien van AI-technologie noodzakelijk. Op dit moment we weten nog onvoldoende wat **verantwoorde toepassing van AI ter ondersteuning van medische beslissingen** inhoudt, volgens relevante stakeholders.

Het doel van dit project was daarom om zorgen, behoeften en wensen van zorgverleners en patiënten ten aanzien van toepassing van AI ter ondersteuning van medische beslissingen in kaart te brengen, om vervolgens de verkregen inzichten te vertalen naar randvoorwaarden voor verantwoorde toepassing van AI.

## METHODE

Om inzicht te krijgen in zorgen, behoeften en wensen ten aanzien van toepassing van AI voor het berekenen van kansen op overleving van IC-opname en implementatie van een dergelijk voorspellend AI-model in de dagelijkse praktijk, hebben we in het 4<sup>de</sup> kwartaal van 2018 **semigestructureerde interviews** uitgevoerd, met aansluitend een **focusgroep bijeenkomst** in februari 2019. Voor de interviews hebben we IC-artsen uitgenodigd, met wie de afdeling Klinische Informatiekunde van Amsterdam UMC onderzoek doet of heeft gedaan naar voorspellende modellen op de IC. We hebben bewust gekozen voor **IC-artsen** die ervaring hebben met onderzoek naar voorspellende modellen, zodat de te ontwikkelen randvoorwaarden op basis van expert opinion geformuleerd zouden kunnen worden. **IC-patiënten of hun naasten** waren ervaren IC-patiëntenvertegenwoordigers vanuit patiëntenorganisatie [IC-connect](#) en vanuit lotgenoten platform [Sepsis en daarna](#). Naast IC-artsen en voormalige IC-patiënten of hun naasten, hebben we ook **technologie experts** uitgenodigd voor een interview en de focusgroep bijeenkomst. Door zowel de eindgebruikers (IC-artsen en IC-patiënten of naasten) als ontwikkelaars van toekomstige voorspellende AI-modellen (technologie experts) uit te nodigen, hebben we getracht om een breed perspectief in beeld te krijgen en daarmee inhoud te geven aan wat verantwoorde AI is volgens **relevante stakeholders**.

De vragen die geïnccludeerd zijn in de semigestructureerde interviews zijn ontworpen op basis van een **literatuurstudie** en drie **verkennende gesprekken** met de volgende experts:

- Hans van Delden – hoogleraar medische ethiek UMC Utrecht,
- Paulien Melis - programma developer Creative Care Lab van Stichting WAAG,
- Sabine Wildevuur – programmaleider Creative Care Lab van Stichting WAAG,

- Aranka Akkermans & Maartje Harmelink – PhD studenten medische psychologie en medische communicatie, betrokken bij een wetenschappelijke studie naar de besluitvorming op IC.

Alle voorbereidende gesprekken en interviews zijn uitgevoerd door projectleider Joanna Klopotowska en onderzoeksassistente Irene Stellingwerf. Van de interviews zijn met toestemming van de deelnemers geluidsopnamen gemaakt, die na afloop woordelijk getranscribeerd zijn door Irene Stellingwerf. Alle getranscribeerde interviews zijn vervolgens in duplo thematisch gecodeerd door Irene Stellingwerf en Joanna Klopotowska om de belangrijkste thema's te kunnen achterhalen en om meningen van geïnterviewden te kunnen kwantificeren per thema. Tijdens de focusgroepbijeenkomst zijn de belangrijkste bevindingen uit de individuele interviews gedeeld met de aanwezige IC-artsen, voormalige IC-patiënten of naasten en technologie experts. Het doel van de focusgroep bijeenkomst was om de geconstateerde verschillen in standpunten nader te onderzoeken en waar mogelijk consensus te bereiken. De focusgroep bijeenkomst werd begeleid door projectleider Joanna Klopotowska en vertegenwoordiger van [Stichting WAAG](#), Paulien Melis. Het projectteam fungeerde als klankbordgroep en adviesorgaan gedurende de gehele looptijd van het project.

## RESULTATEN

Er zijn in totaal tien IC-artsen (zowel van academische als niet-academische ziekenhuizen), vijf IC-patiëntvertegenwoordigers (waarvan vier voormalige IC-patiënten en één naaste van een voormalig IC-patiënt) en twee technologie experts geïnterviewd. Beide technologie experts hebben medische informatiekunde als achtergrond. Tijdens de focusgroep bijeenkomst waren er vier van de tien geïnterviewde IC-artsen aanwezig, drie van de vijf IC-patiëntvertegenwoordigers, en één van de twee technologie experts. Behalve de geïnterviewden namen projectteam leden, twee vertegenwoordigers van [Pacmed](#) en één senior onderzoeker van [Tilburg Instituut voor Law, Technology en Society \(TILT\)](#) deel aan de discussie.

De informatie uit de interviews en de focusgroep bijeenkomst is vertaald naar een **set van 22 randvoorwaarden** voor een verantwoorde toepassing van AI-technologie ter ondersteuning van medische beslissingen via voorspellende AI-modellen. Deze randvoorwaarden **weerspiegelen zorgen, behoeften en wensen van het merendeel van zorgverleners, patiënten en technologie experts** die deel hebben genomen aan dit onderzoek. Indien van toepassing zijn de opgestelde randvoorwaarden voorzien van extra toelichting uit literatuuronderzoek. Veel van de randvoorwaarden zijn niet AI-specifiek, maar gelden ook voor simpele voorspellende modellen. Ook zonder AI valt er namelijk nog veel winst te behalen als het gaat om beslissingsondersteuning door computers in het medisch domein. Echter, door toepassing van AI worden voorspellende modellen preciezer en daarmee wordt het stellen van concrete randvoorwaarden belangrijker. De hieronder gepresenteerde set van 22 randvoorwaarden is ingedeeld in **6 thematische groepen**, waarbij per thema enkele quotes van de deelnemers zijn uitgelicht:

1. Randvoorwaarden ten aanzien van **welke medische beslissingen** te ondersteunen met voorspellende AI-modellen (blz. 12).
2. Randvoorwaarden ten aanzien van **(medische) gegevens** die gebruikt worden in voorspellende AI-modellen (blz. 14).
3. Randvoorwaarden ten aanzien van **ontwikkelaars** van voorspellende AI-modellen (blz. 19).
4. Randvoorwaarden ten aanzien van het **ontwikkeling en evaluatie** van voorspellende AI-modellen (blz. 24).
5. Randvoorwaarden ten aanzien van **gebruik** voorspellende AI-modellen in de klinische praktijk (blz. 28).
6. Randvoorwaarden ten aanzien van **communicatie uitkomsten** voorspellende AI-modellen aan de patiënt (blz. 32).

## 1. RANDVOORWAARDEN TEN AANZIEN VAN WELKE MEDISCHE BESLISSINGEN TE ONDERSTEUNEN MET VOORSPELENDE AI-MODELLEN

*“Het blijkt uit onderzoek dat artsen vaak te optimistisch zijn over de overlevingskansen en de verpleegkundigen te somber. Welke kant je ook kiest, het gevolg is dat je soms patiënten onnodig doorbehandeld en soms dat je patiënten ten onrechte abstineert. Dus er is behoefte aan informatie waarmee je een betere prognose kunt stellen. Dat is belangrijk voor de patiënt of in dit geval meestal voor de naasten.”*

*“Ik zou artificiële intelligentie in eerste instantie niet inzetten voor het voorspellen van kansen op overleving op de Intensive Care. Dat laat zich niet gemakkelijk in een getal uitdrukken. Mijn advies zou zijn, begin maar eerst bij wat harde meetbare zaken. In de toekomst als men het meer gewend is en het werkt dan kun je het gaan inzetten voor meer complexe problemen.”*

*“Ik ben nu misschien met allemaal zaken rekening aan het houden die totaal geen invloed hebben op de uitkomst, omdat ik er heel erg van overtuigd ben dat dat wel zo is. Je wilt ook verrast worden om nieuwe aannames te moeten doen. Nieuwe relaties die je zelf nooit had gelegd, zodat je daar weer onderzoek naar kunt doen.”*

*“Het gaat nu vooral op ervaring van de artsen....Ik zou wel meer onderbouwing willen hebben. Ik zou onderbouwing willen hebben, of de arts overlegd heeft met anderen, waarop is de prognose gebaseerd, dat soort vragen. Het gaat nu ook vaak even snel tussendoor dat je informatie krijgt....Wat fijner zou zijn als je ergens rustig zou kunnen overleggen met de familie met ook speciale begeleiding.”*

*“Zo een model dat is alleen maar prima, hoe meer informatie, hoe beter de artsen naar iets kunnen refereren. Maar persoonlijke aandacht voor de mensen is het meest ideale.”*

*“Ik heb verschillende ervaringen met informatieverstrekking op de IC en ook luisteren naar de inbreng van de familie. Op de ene IC werd ik pas na veel aandringen meegenomen in de gesprekken. Op een andere IC waren artsen ook eerlijk over dat er heel veel onzekerheid was, maar ze hadden wel steeds een vervolgplan.”*

## 1.1. Voorspellende AI-modellen moeten voorzien in behoeften van hun eindgebruikers.

Bij het nemen van medische beslissingen worden **diverse informatiebronnen** geraadpleegd en onderzoeken uitgevoerd. Hoe complexer een vraagstuk, hoe meer informatiebronnen en onderzoeken geraadpleegd worden. En dan nog is in veel gevallen niet direct sprake van een diagnose of beslissing, maar worden diverse hypothesen opgesteld en getoetst. Dat levert een bepaalde mate van waarschijnlijkheid (een kans) op dat het om uitkomst A gaat en niet om uitkomst B of C. De ervaring, expertise en vaak ook intuïtie van een behandelteam is van grote invloed op het inschatten van kansen. Bij **complexe vraagstukken** is daarom gebruikelijk om zogenaamde multidisciplinaire overleggen te organiseren. Bij zo'n overleg zijn meerdere experts van **diverse specialismen** aanwezig. Soms worden ook naasten van een patiënt uitgenodigd om deel te nemen of, indien mogelijk, patiënten zelf. Niet zelden spelen culturele en organisatorische normen en waarden ook een rol bij het nemen van dergelijke complexe beslissingen.

Tegelijkertijd heerst bij veel artsen regelmatig het gevoel door “de bomen het bos niet meer te zien”. De verregaande digitalisering in de zorg heeft bijgedragen aan **exponentieel groei van (medische) gegevens**. Daarnaast worden elke dag tientallen wetenschappelijke artikelen gepubliceerd met nieuwe inzichten over het ontstaan van ziekten en hun behandeling. Echter, de **tijd** die de zorgverleners hebben per patiënt is helaas steeds korter geworden, wat een diepgaande analyse van alle beschikbare informatie vaak in de weg staat. De meerwaarde van voorspellende AI-modellen ligt volgens de artsen daarom vooral daar waar grote hoeveelheid diverse gegevens in een relatief korte tijd verwerkt moeten worden. Hierbij hopen de zorgverleners dat voorspellende AI-modellen hen helpen om **beter gefundeerde beslissingen** te nemen, waar nu te veel geleund wordt op “gut feeling”. Ook verwacht men ondersteuning bij het **herkennen van patronen** die zorgverleners zelf niet snel genoeg zouden herkennen en bij het vinden van nieuwe **aanwijzingen voor nader onderzoek**. Aan de andere kant is men ook terughoudend als het gaat om het inzetten van voorspellende AI-modellen ter ondersteuning van complexe beslissingen, zoals in het geval van een voorspellend model om kansen op overleving van IC-opname te berekenen. Dit heeft vooral te maken met het **gebrek aan betrouwbare gegevens** over bijvoorbeeld het functioneren van de patiënt voor en na de IC-opname (zie paragraaf 2.3). Dit zijn gegevens die volgens artsen en patiënten van groot belang zijn om een goede inschatting van kansen te kunnen maken. Andere verklaringen zouden kunnen zijn: **onvoldoende bekendheid met AI-technologie** en **gebrek aan wetenschappelijk bewijs** voor meerwaarde van voorspellende AI-modellen en, daarmee samenhangend, gebrek aan **succesvolle toepassing in de klinische praktijk**.

Patiënten waren overwegend positief over de inzet van AI voor het ontwikkelen van voorspellende modellen ter ondersteuning van een complexe beslissing, zoals wel of niet doorgaan met levensbehoudende IC-behandeling. Zij en/of hun naasten hebben regelmatig situaties meegemaakt waarin veel onzekerheid heerste over de kans op IC-overleving. Elke extra informatie zien patiënten dus als toegevoegde waarde. Patiënten gaven echter ook aan dat hun behoefte aan duidelijke **communicatie en persoonlijke aandacht** eigenlijk nog belangrijker zijn. Dit betreft behoeften die voorspellende AI-modellen niet kunnen invullen en waar zorgverleners aandacht voor moeten hebben. Juist wanneer het gevaar bestaat dat patiënten en naasten kunnen denken dat een computer, en niet de behandelend arts, een beslissing neemt over iets delicaats als levensverlengende behandeling.

Een gedegen **context analyse** voorafgaand aan de ontwikkeling van een voorspellend model samen met alle betrokken stakeholders, zou daarom de eerste stap moeten zijn om alle facetten van een medische beslissing die een voorspellend model zou moeten ondersteunen in kaart te brengen. Specifiek voor voorspellende AI-modellen, dient vooral de gewenste mate van betrouwbaarheid van de risicoschatting van de uitkomst waarover een beslissing moet worden genomen nader onderzocht te worden. Een context analyse zou kunnen helpen om **realistische verwachtingen** te schetsen of een voorspellend model wel of niet zou kunnen voorzien in de behoeften en wensen van de betrokken stakeholders en daarmee daadwerkelijk gebruikt zou gaan worden ter ondersteuning van medische beslissingen in de klinische praktijk.



## 2. RANDVOORWAARDEN TEN AANZIEN VAN (MEDISCHE) GEGEVENS DIE GEBRUIKT WORDEN IN VOORSPELENDE AI-MODELLEN

*“Zo’n model dat allerlei data uit ons EPD gebruikt en ondertussen weet ik dat daarin veel registratiefouten zitten...ik zou toch best huiverig zijn voor onjuiste prognoses voor individuele patiënten.”*

*“Voordat ik zo’n model ga gebruiken, wil ik weten wat de beperkingen ervan zijn, dus bijv. welke patiëntengroepen zijn onvoldoende vertegenwoordigd in de data waarop een model zijn voorspelling baseert.”*

*“We missen nu heel vaak een goed overzicht van de voorgeschiedenis, wat volgens mij van groot belang is voor de prognose. We hebben dat vaak wel, maar niet in een soort bruikbaar pakket voor dataverwerking. Is eigenlijk een administratief probleem.”*

*“Er is nu nauwelijks zicht op hoe het met de patiënt gaat na het ontslag van de Intensive Care (IC) afdeling. Zijn onze medische doelen behaald? Is de patiënt tevreden? Vond de patiënt het de moeite waard? Ook weten we vaak te weinig over hoe de patiënt functioneerde voor die bij ons kwam. Informatie over beloop voor en na de IC-opname zou ons helpen om beter in te schatten wat nog wel en wat niet zinvol is qua behandeling bij nieuwe IC opnamen.”*

*“Zo’n voorspellend AI-model moet beschikken over data die de patiënt meest optimaal beschrijven. Dus wat ik als arts in mijn hoofd doe, dat moet de computer ook weten. Als de data die de computer gebruikt te beperkt is, dan doe ik waarschijnlijk zelf beter.”*

*“Het liefst zou je ook informatie hebben over hoe een patiënt in het leven stond voor de IC-opname. Het is anders heel moeilijk om te vragen aan naasten, want die hebben vaak een andere inschatting. Zeker als mensen al beperkt zijn...want je past je kwaliteit van leven aan je beperking. Dus dat is moeilijk om in te schatten voor de mensen eromheen. En je kunt het vaak niet aan de patiënt zelf vragen.”*

*“Ik denk dat het van groot belang kan zijn om zo veel mogelijk gegevens te hebben om goed werkende modellen te ontwikkelen...maar je moet je altijd afvragen hoeveel al deze gegevens bijdragen aan betere voorspelling van een bepaalde uitkomst versus het inboeten op privacy van de patiënt.”*

*“Er zit zeker bias in EPD-gegevens. Daar moeten we niet lullig over doen. Als een patiënt van 91 op de IC wordt opgenomen dan is de kans groot dat wij ten eerste zo een patiënt niet opnemen en ten tweede dat we dat niet doen voor een langdurige IC opname. En misschien ook sneller de handdoek in de ring gooien dan bij iemand van 30.”*

*“Ik denk dat succes van iedere behandeling voor een groot deel afhangt van factoren die we nu nog niet standaard meten...waarschijnlijk genetica. Maar hier zit wel een ethisch dilemma onder: willen we echt modellen maken waarin ons genetisch profiel meegenomen wordt?”*

*“De meeste artsen kennen de persoon niet die in dat bed ligt, familie wel. Zeker als iemand in een kritieke toestand binnenkomt dan heb je die informatie niet in het EPD.”*

## 2.1. Gegevens zijn van voldoende kwaliteit.

De geïnterviewden gaven aan dat het van cruciaal belang is dat de gegevens die gebruikt worden door voorspellende AI-modellen van voldoende kwaliteit zijn. De kwaliteit van de gebruikte gegevens bepaalt immers voor het grootste deel de kwaliteit van de uitkomst van een voorspellend AI-model. **“Garbage in, is garbage out.”**

Er bestaan verschillende definities van **datakwaliteit**. Vaak is datakwaliteit gedefinieerd in termen van “de dimensies van datakwaliteit”. De meeste definities omvatten de volgende aspecten:

- Compleetheid: alle bekende data van een casus en alle casussen zijn geregistreerd,
- Uniceit: de data komt niet dubbel voor,
- Validiteit: de waarden meten wat wordt bedoeld,
- Juistheid: de data komt overeen met de werkelijkheid,
- Tijdigheid: de data is niet te oud om relevant te zijn voor het doel waarvoor ze gebruikt wordt.

In het algemeen worden de gegevens beschouwd als “gegevens met een hoge kwaliteit” als ze geschikt zijn voor het beoogde gebruik. Met andere woorden, de eisen aan kwaliteit van data zijn afhankelijk van het beoogde doel. Gegevens die dagelijks door mensen en apparaten gegenereerd worden tijdens het zorgproces en al dan niet geregistreerd worden in **het elektronisch patiëntendossier (EPD)** zijn onderhevig aan diverse risico's voor datakwaliteit. Dit heeft te maken met het feit met het primaire doel van gegevensverzameling via het EPD: de registratie ten behoeve van zorgverlening (en niet gegevensanalyse). Drie belangrijke risico's zijn: dataverlies, fouten en verkeerd gebruik van de systemen waar de gegevens worden ingevoerd.

Bij **dataverlies** moet men denken aan situaties zoals het onvermogen om gegevens in te voeren als gevolg van het uitvallen van diensten en services, zoals bijvoorbeeld stroomuitval of netwerkstoringen en dataverlies door het verlies van opgeslagen gegevens.

**Fouten** kunnen geclassificeerd worden in twee categorieën: data-entry fouten en transcriptiefouten. Data-entry fouten zijn problemen zoals typerfouten, een selectie van een verkeerd item uit een keuzelijst, klikken op een verkeerde knop of fouten veroorzaakt door onjuist begrip van een invulveld. Transcriptiefouten kunnen ontstaan als eindgebruikers gegevens moeten overtypen, bijvoorbeeld vanuit een EPD naar een ander systeem, zoals bij kwaliteitsregistraties.

Bij **verkeerd gebruik van de systemen** gaat het vooral om problemen waarbij de benodigde gegevens wel zijn geregistreerd, maar niet op de juiste plek of in het juiste formaat. Veel artsen willen liever gegevens registreren door het invoeren van een "verhaal" in vorm van een vrije tekst, dan via het invoeren van gestructureerde/gestandaardiseerde gegevens. Zo zien we vaak dat diagnoses als vrije tekst in een klinische notities worden geregistreerd, in plaats van als een gestructureerde diagnose code in een diagnose keuzeveld. Het gestructureerd vastleggen van diagnose codes is onder andere nodig om beslissingsondersteuning te kunnen bieden, bijvoorbeeld over welke therapie het beste past bij een diagnose. Onder verkeerd gebruik vallen ook problemen die veroorzaakt worden door zogenaamde workarounds. Indien in het EPD geen passende functionaliteit bestaat voor invoer van vereiste gegevens, of als het moeilijk is om het juiste veld te vinden, dan bedenken zorgverleners meestal een eigen oplossing, een workaround. Bijvoorbeeld: Een verpleegkundige dient eerst alle medicatie aan de patiënten toe en registreert alle toedieningen achteraf in het EPD om tijd te besparen. De gegevens staan wel in het systeem, maar het tijdstip van de toedieningen in het EPD komt niet overeenkomt met het daadwerkelijke tijdstip van de toedieningen, waardoor in analyses verkeerde conclusies getrokken zouden kunnen worden.

Wat betreft **compleetheid van data** kan zowel onderregistratie als overregistratie voorkomen. Zo kunnen bepaalde activiteiten (datapunten) niet geregistreerd worden die wel zijn uitgevoerd, bijvoorbeeld: het meten van een pijnscore bij een patiënt zonder de score vast te leggen in het EPD. Er kunnen ook bepaalde activiteiten geregistreerd worden die niet zijn uitgevoerd, bijvoorbeeld: een activiteit wordt vooraf als uitgevoerd geregistreerd, maar op het betreffende moment is de

conditie van een patiënt verslechterd en wordt de activiteit niet meer uitgevoerd, maar niet verwijderd uit het EPD.

Om deze bovengenoemde redenen is het hergebruiken van routinematig ingevoerde medische gegevens uit bijv. het EPD voor het ontwikkelen van voorspellende AI-modellen niet goed mogelijk zonder een **stevige dosis aan controles, validatie en transformaties**. Omdat in elk ziekenhuis het EPD vaak anders is ingericht en op een andere manier wordt gebruikt, moeten deze acties meestal per ziekenhuis en per dataset herhaald worden. Het transparant zijn naar de eindgebruikers over de datakwaliteit, de uitgevoerde controles, validatie en transformaties en daaruit voortgekomen bevindingen, is van belang om de **bependingen van de ontworpen voorspellende AI-modellen** goed te kunnen begrijpen.

## 2.2. Gegevens bieden voldoende afspiegeling van de diversiteit aan patiënten/settings.

Elk ziekenhuis heeft zijn eigen patiënten, een eigen profiel, bijvoorbeeld in een academisch ziekenhuis worden vaker patiënten met zeldzame aandoening behandeld in vergelijking met een niet-academisch ziekenhuis. En ook de **cultuur in de regio en cultuur op een bepaalde ziekenhuisafdeling** kunnen van invloed zijn op het medisch beleid. Zo kunnen door bepaalde culturele of religieuze invloeden verschillen ontstaan in de duur van levensbehoudende IC-behandelingen, omdat er verschillend gedacht wordt over levensbeëindiging. Verder kan representatie van minderheden (bijvoorbeeld op grond van socio-economische status, etniciteit, religie, leeftijd) in EPD-data per ziekenhuis verschillen. Bij te lage aantallen van bepaalde groepen patiënten kunnen geen betrouwbare voorspellingen voor deze groepen gedaan worden via machine learning. Om deze redenen kun je een voorspellend AI-model, ontwikkeld op basis van gegevens uit één ziekenhuis, niet zondermeer overnemen en toepassen in een ander ziekenhuis. Het verdient dus de voorkeur om gegevens uit **diverse ziekenhuizen** te combineren bij de ontwikkeling van voorspellende AI-modellen en te zorgen dat voldoende aantallen van minderheden gerepresenteerd worden in een te bestuderen dataset. Het **samenwerken** met andere centra biedt bovendien een mooie gelegenheid om met elkaar in discussie te gaan over de gevonden verschillen in uitkomsten (indien van toepassing) en daarvan te **leren**. Uiteraard kunnen er specifieke lokale behoeften zijn om AI-technologie in te zetten waardoor men zou kunnen volstaan met het gebruik van data uit eigen setting en evaluatie van het model op deze data.

## 2.3. Gegevens zijn meetbaar en in voldoende mate en tijdig beschikbaar.

Voordat men aan de slag gaat met het ontwikkelen van voorspellende AI-modellen met behulp van AI, is het van belang om goed in kaart te brengen **welke informatie en dus welke gegevens noodzakelijk zijn** om een model te kunnen ontwerpen van klinische meerwaarde. In het geval van een voorspellend AI-model voor kans op overleving van een IC-opname, bleek dat naast gegevens beschikbaar in het **EPD** ook gegevens **buiten de medische context** van belang zijn om een goede inschatting te kunnen maken of het wel of niet zinvol is om door te gaan met levensbehoudende IC-behandeling. Deze gegevens zijn op dit moment onvoldoende of zelfs niet beschikbaar. Bijvoorbeeld informatie over het functioneren van de patiënten voor de IC-opname: *“Wat wil een patiënt zelf?”, “Wat kon een patiënt zelf?”, “Wat was de conditie?”, “Welke doelen zijn haalbaar voor een patiënt?”* en informatie over het functioneren van de patiënten die de IC levend hebben verlaten: *“Welke kwaliteit van leven ervaart de patiënt”, “Wat lukt nog wel en wat lukt niet meer?”*. Een complicerende factor in het geval van beslissing om een levensbehoudende IC-behandeling wel of niet te staken, is dat de patiënt zelf meestal niet antwoord kan geven op deze vragen en de naasten mogelijk onvoldoende in staat zijn om een goede inschatting te maken wat voor de patiënt aanvaardbaar is.

Daarnaast zijn er aspecten van het functioneren van de patiënten die zich **niet zondermeer in een getal laten uitdrukken**, zoals de wil tot leven, levenslust of geluksgevoel. Zelfs in het geval van kwaliteit van leven waarvoor diverse scoringsinstrumenten beschikbaar zijn, blijft het lastig om een conclusie aan een score te verbinden, omdat een score niet per se een betrouwbare weergave is van de waarde die de patiënt aan zijn of haar kwaliteit van leven toekent. Ook blijken mensen over een enorm aanpassingsvermogen te beschikken als het gaat om het omgaan met



beperkingen ontstaan door ziekte. Zo kunnen patiënten ondanks een beperking alsnog een goede kwaliteit van leven ervaren, terwijl ze dat als onmogelijk achtten voordat deze beperking ontstond. Verder dienen de **benodigde gegevens beschikbaar te zijn op het moment van besluitvorming**. Indien in het model gegevens gebruikt worden die verzameld zijn nadat een behandelteam tot een bepaald besluit is gekomen, dan zal een voorspelling door zo'n model (die "achteraf" gegevens meeneemt) van weinig meerwaarde zijn voor de dagelijkse praktijk. Eén van de experts geïnterviewd in dit project vertelde over een voorspellend model dat heel goed in staat was om te voorspellen dat een patiënt ging overlijden. Echter, dat kwam niet door de slimheid van het model, maar door het soort gegevens dat in het model werden gebruikt. Dit model gebruikte namelijk gegevens die heel duidelijke tekens waren dat het stervensproces zich aan het voltrekken was, zoals gegevens die wijzen op vergevorderde multiple orgaan falen. Het model kwam dus te laat met een voorspelling en voegde niks toe om artsen bij hun beslissing te ondersteunen.

Een ander belangrijke voorwaarde is dat er voldoende gegevens en meetpunten beschikbaar zijn in de data voor de groep patiënten die men wil bestuderen. Zo blijkt uit onderzoek beschreven in het artikel van Gainfrancesco et al. dat personen uit kwetsbare populaties zoals personen met lage socio-economische status, personen met psychosociale problemen en immigranten, vaker verschillende zorginstellingen bezoeken en bijvoorbeeld minder vaak gegevens invoeren via online patiëntenportalen. Hierdoor raakt de informatie over deze groepen patiënten versnipperd en/of is helemaal niet beschikbaar, waardoor vertekening kan ontstaan in voorspellingen. Als informatie over een bepaalde groep patiënten in onvoldoende mate aanwezig is, kan een voorspellend AI-model namelijk geen betrouwbare uitkomsten toekennen aan deze patiënten. Om dit probleem te beperken, kunnen ontwikkelaars verschillende technieken gebruiken om op de juiste manier om te gaan met ondervertegenwoordiging van bepaalde groepen in de gegevens. Dit gebeurt echter nog onvoldoende.

#### **2.4. Privacy van de patiënten is voldoende geborgd.**

Als je in het ziekenhuis opgenomen wordt, dan geef je impliciet een **toestemming** om allerlei gegevens over je te verzamelen die nodig zijn voor je behandeling. Echter, als men deze **gegevens** vervolgens wil **gebruiken voor andere doeleinden**, bijv. analyses in het kader van wetenschappelijk onderzoek, dan is een expliciete toestemming van de patiënten nodig. Tenzij de gegevens **geanonimiseerd** worden en in redelijkheid niet herleidbaar gemaakt kunnen worden tot één persoon of een kleine groep personen. Echter, zeker als het gaat om big-data, is het anonimiseren van medische datasets niet goed mogelijk, omdat de hoeveelheid aan diverse gegevens per patiënt zo groot is dat er toch een kans bestaat om alsnog een persoon of kleine groepen personen aan te kunnen wijzen. Daarom is er meestal sprake van **gepseudonimiseerde medische gegevens**. Dit betekent dat alle gegevens herleidbaar tot één persoon of een kleine groep personen zoals naam, geboortedatum, adres, BSN en patiëntnummer uit de dataset verwijderd zijn. Elke patiënt of een opname krijgt vervolgens een unieke code die geen tot persoon herleidbare informatie bevat. De sleutelrijst, oftewel de link tussen zo'n unieke code en de herleidbare gegevens, wordt meestal door een trusted third party bewaard. Bij gepseudonimiseerde medische datasets is onder enkele voorwaarden (bijvoorbeeld toestemming vragen is redelijkerwijs niet mogelijk door grote aantallen of type patiënt) geen expliciete toestemming van de patiënten nodig, indien men deze gegevens voor wetenschappelijk onderzoek wil gebruiken.

Echter, als het gaat om onze gezondheid en bijv. het voorspellen van kansen op ziekte en herstel met behulp van AI-modellen, dan zijn **gegevens uit ons dagelijks leven** (dus buiten de medische context), ook interessant om mee te nemen bij de ontwikkeling van voorspellende AI-modellen. Zoals in het geval van het functioneren van een patiënt voor en na de IC-opname. Dit zouden gegevens kunnen zijn die verzameld worden via Google, Facebook en LinkedIn over o.a. je sociale context, leefstijl en opleiding, via gezondheidsapps over o.a. mate van bewegen, of via Albert Heijn app over je voedingspatroon. Door het **samenvoegen van diverse datasets** wordt het **waarborgen van privacy een stuk moeilijker**. De Algemene Verordening Persoonsgegevens (AVG) die in 2018 in de Europese Unie is ingevoerd, biedt deels bescherming tegen privacy inbreuk, maar het is juist de combinatie van diverse al dan niet gepseudonimiseerde of zelfs geanonimiseerde data sets dat maakt dat het veel makkelijker wordt om één persoon of groep van

personen te identificeren. Nog complexer wordt het op het moment dat ook **genetische informatie** over individuen als databron voor de ontwikkeling van voorspellende AI-modellen wordt gebruikt. Immers, genetische informatie is direct identificeerbaar tot één persoon.

Een mogelijke oplossing zou kunnen zijn: het ontwikkelen van decentraal leren. In deze situatie blijven de gegevensverzamelingen lokaal (en moet de betekenis, bijvoorbeeld het gebruikte terminologiestelsel om de gegevens te coderen, van ieder gegeven in de set goed beschreven worden) en wordt het algoritme langs de dataverzamelingen gestuurd om een model te leren zonder dat deze gegevens fysiek de bron verlaten. De zo ontworpen AI-modellen kunnen lokaal uitgevoerd worden en de uitkomsten gedeeld worden.

## **2.5. Klinische redentatie moet uit de gegevens af te leiden zijn.**

Het wel of niet uitvoeren van diagnostisch onderzoek zoals MRI of CT-scans, monitoring van klinisch chemische parameters zoals kalium of hemoglobine, of het voorschrijven van geneesmiddelen gebeurt meestal in de lijn van **klinische redentatie van een behandelteam** over wat de patiënt mankeert en hoe ernstig de situatie is. Oftewel, wat in het EPD geregistreerd staat of juist niet geregistreerd staat is een **reflectie van het beleid**, van de keuzes die in een ziekenhuis gemaakt zijn. Dit betekent dat de hoeveelheid en soort informatie niet voor alle patiënten in dezelfde mate of überhaupt niet terug te vinden is in het EPD. Er is dus mogelijk sprake van een **vertekening of bias** in dergelijke gegevens. Vertekening die volgt uit menselijke redentatie van de betrokken zorgverleners. Bijvoorbeeld: als een behandelteam over het algemeen van mening is dat een IC-behandeling niet meer zinvol is bij patiënten ouder dan 90 jaar, dan zal de frequentie van diagnostiek en monitoring afnemen en/of zullen specifieke activiteiten geregistreerd worden die richting palliatief beleid wijzen. Gegevens die zouden kunnen wijzen op een tegenovergesteld situatie, dus dat het wel zinvol is om IC-behandeling voort te zetten, zijn dan niet beschikbaar. Een voorspelend AI-model dat op basis van gegevens van een zo'n IC-afdeling ontwikkeld wordt, zal daarom een lagere kans op overleving afgeven voor patiënten ouder dan 90 jaar. Op die manier kan een AI-model tot een zelf vervullende voorspelling of **self-fulfilling prophecy** leiden. De socioloog Robert K. Merton definieerde het begrip als volgt:

*"De zelf vervullende voorspelling is in aanvang een foute definitie van de situatie die een nieuw gedrag oproept waardoor de oorspronkelijke foute kijk waar wordt. Deze schijnbare juistheid van de voorspelling houdt een foute voorstelling van zaken in stand. De voorspeller zal namelijk datgene wat uiteindelijk gebeurd is aanvoeren als bewijs dat hij van begin af aan gelijk had."*

Daarom is het van belang om naast alle meetbare gegevens ook de achterliggende klinische redentatie te weten en daarmee dus ook te weten welke andere mogelijke behandelopties niet onderzocht zijn. Alleen dan kun je betekenis verbinden aan wat je leert op basis van de gegevens die je krijgt. Helaas wordt de achterliggende klinische redentatie vaak niet in het EPD geregistreerd. Een manier om met deze beperking in data om te gaan en daarmee mogelijke vertekeningen op te sporen, is het **combineren van gegevens uit EPDs van diverse zorgaanbieders**. Op die manier kunnen patronen in behandelingen van één zorgaanbieder vergeleken worden met patronen die men leert op basis van data bij andere zorgaanbieders. Het verschil in dergelijke patronen zou kunnen wijzen op bias. Bijvoorbeeld, als het blijkt dat op een IC-afdeling patiënten ouder dan 90 jaar een lage kans hebben op overleving van IC-opname hebben maar deze kansen veel hoger zijn op andere IC-afdelingen, dan kan dit dus wijzen op een behandelpatroon waarin leeftijd het klinisch redeneren in belangrijke mate stuurt ten aanzien van besluitvorming over wel of niet doorgaan met IC-behandeling.

### 3. RANDVOORWAARDEN TEN AANZIEN VAN ONTWIKKELAARS VAN VOORSPELENDE AI-MODELLEN

*"In zo een team moeten mensen zitten die precies weten welke fouten in de data aanwezig kunnen zijn, fouten in het programmeren kunnen opsporen, maar ook weten: hier zit helemaal niemand op te wachten."*

*"AI-experts die bij Google werken die zouden prima zo een model kunnen ontwikkelen, maar ik vind het geen geschikte partij voor in de zorg. Commerciële partijen hebben nu eenmaal andere belangen en die zijn ook niet zo zuinig op privacy, vind ik."*

*"Ik vind zorgverzekeraars geen goede partner bij de ontwikkeling van dergelijke modellen. Stel je voor dat ze vervolgens een IC opname niet vergoeden, omdat op basis van een model iemand in een bepaald profiel past."*

*"Modellen zijn niet waarde vrij, want in de data zitten onze keuzes, wat we wel of niet toelieten in de behandeling. En we zijn niet waarde vrij."*

*"Qua intellectueel eigendom van zo een model is dat hier op de universiteit heel simpel...dan is het van de universiteit. Er zijn natuurlijk commerciële bedrijven die ook AI-modellen maken en dan geldt voor vragen en het model niet openbaar maken. Vind ik niet van deze tijd."*

*Qua intellectueel eigendom denk ik aan drie dingen: je hebt de data, het systeem en trainingstijd. Als je het hebt over machine learning, dan is trainingstijd een heel groot gedeelte van de investering. Je zou kunnen zeggen, die trainingstijd is betaald door de burgers die via belastingen betalen voor het onderwijs van artsen die de training doen. Het is een interessante puzzel."*

*"Wat betreft verantwoordelijkheid dan denk ik dat je er vrij makkelijk een disclaimer voor kan schrijven. Dat het hier gemaakt is, op basis van onze data, dat er uiteraard grote mate van onbetrouwbaarheid in zit. Dat het hoogstens een ondersteuning kan zijn bij besluitvorming die op basis van veel verschillende componenten tot stand komt. Denk dat je dat op die manier wel kan ondervangen."*

*"Diegene die zo een model distribueert, is wat mij betreft verantwoordelijk voor het testen van zo een model in de klinische praktijk. Zo een model zal niet perfect zijn, dat moeten we accepteren."*

*"Voorlopig blijft de dokter nog eindverantwoordelijke. Wij kiezen ervoor om zo een technologie te gebruiken. Wij moeten ons daarom ervan vergewissen dat het veilig is voordat we het gaan inzetten."*

*"Stel je hebt een model om te voorspellen of het zinvol is om iemand een schaarse en dure therapie te geven. De computer zegt: hier ligt de grens van de kans, jij krijgt het wel of niet. Dat wordt het wel interessant als de patiënt zegt: heb ik daar niet gewoon recht op? Klopt jouw model wel? Dat zijn wel spannende vragen en daar is nog geen duidelijkheid over. Waar ligt die verantwoordelijkheid eigenlijk? Wat als je ernaast zit? En natuurlijk moet je zo een model goed valideren, maar het wordt nooit perfect."*

*"Het is echt ontzettend belangrijk om te weten wie een model heeft getraind en met welke data."*

### 3.1. Voorspellende AI-modellen worden ontwikkeld door een multidisciplinair team.

Het ontwikkelen van voorspellende AI-modellen om medische beslissingen te ondersteunen moet een resultaat zijn van **teamwork**. In zo'n team participeren experts die een achtergrond hebben in medische informatiekunde, statistiek, data-science, experts die het klinisch (data) domein goed kennen, experts die kunnen programmeren, en potentiële eindgebruikers zoals artsen, verpleegkundigen en patiënten. Afhankelijk van de beoogde toepassing van een voorspellend AI-model zal participatie van een medisch ethicus, medisch psycholoog en communicatiedeskundigen sterk overwogen dienen te worden.

**Participatie van zorgverzekeraars en commerciële partijen** bij de ontwikkeling van voorspellende AI-modellen om kansen op overleving van IC-opname te voorspellen, wordt als risicovol beschouwd. De overtuiging dat de **belangen en doelstellingen** van deze partijen niet op één lijn liggen met die van zorgverleners en patiënten, maakt dat er sprake is van **laag vertrouwen** in deze partijen. Zo denkt men dat het belang van zorgverzekeraar om zorg betaalbaar te houden, mogelijk tot voorspellende AI-modellen zou kunnen leiden waarin kosten van IC-zorg sterker mee gewogen worden dan wensen van de patiënt en de naasten. Als het gaat om samenwerking met commerciële partijen, dan is men er vooral huiverig voor dat de ontwikkeling van voorspellende AI-modellen met of door zo'n partij minder zorgvuldig zou plaatsvinden omwille van tijd en geld.

In een recent artikel van Brent Mittelstadt van Oxford Internet Institute en Alan Turing Institute wordt **gebrek aan gemeenschappelijke doelstellingen en fiduciaire verplichtingen** van partijen die zich met ontwikkeling van voorspellende AI-modellen bezig houden als één van de belangrijkste risico's genoemd voor verantwoorde ontwikkeling en toepassing van AI. Fiduciaire verplichtingen zijn verplichtingen tot eerlijk en verantwoord handelen in het belang van een andere persoon. Dit is in essentie wat patiënten verwachten van een arts die hun behandelt en deze verplichting is niet alleen een goed voornemen van artsen, maar een medische norm die sterk geformaliseerd is via professionele gedragscodes en juridische kaders. Vergelijkbare verplichting is niet zondermeer gegarandeerd als het gaat om commerciële AI-partijen en complementaire gedragscodes en juridische kaders ontbreken op dit moment. Commerciële partijen en hun medewerkers hebben in eerste plaats fiduciaire verplichtingen jegens hun aandeelhouders. Ondertekenen van (zelf opgestelde) AI-gedragscodes zonder duidelijk omschreven en afdwingbare verplichtingen is goed voor de reputatie van een bedrijf naar buiten toe, maar onvoldoende om te garanderen dat zo'n partij zich ook daadwerkelijk eraan houdt. Uit een meta-analyse van Kish-Gephart et al. naar gedragscodes en hun effect op onethisch gedrag, blijkt dat het ondertekenen van (zelf opgestelde) gedragscodes onethisch gedrag niet voorkomt. Gedragscodes en hun onderliggende principes bleken alleen effectief, indien deze ingebed waren in de organisatiecultuur en er actief op gehandhaafd werd (intern en extern).

Indien samenwerking met zorgverzekeraars en/of commerciële partijen beoogd wordt bij ontwikkeling van voorspellende AI-modellen, is het helder krijgen van de belangen van alle teamleden een manier om de zorgen bij zorgverleners en patiënten weg te nemen. De afstemming van deze belangen dienen ook **contractueel en transparant vast te worden gelegd**. Op bestuurlijk niveau lijken duidelijk geformuleerde eisen ten aanzien van ontwikkeling, evaluatie en implementatie van voorspellende AI-modellen, documentatie van deze stappen en ethische en wetenschappelijke beoordeling en monitoring noodzakelijk.

### 3.2. Ontwikkelaars zijn transparant over de mogelijke bias in gegevens waarop hun voorspellende AI-modellen zijn ontwikkeld.

De keuzes die gemaakt worden tijdens de ontwikkeling van een voorspellend AI-model kunnen via **diverse mechanismen** tot vertekeningen (**bias**) in de voorspellingen/ uitkomsten van voorspellende AI-modellen leiden. De vijf meest bekende mechanismen zijn:

- Het definiëren van uitkomst variabele;
- Keuze van trainingsdata;
- Selectie van kenmerken;
- Onbedoelde proxy's;
- Maskeren van intenties van ontwikkelaars.

Om te leren voorspellen krijgt de computer meestal een zogenaamde trainingsdata of trainingset van historische gegevens, waarin per persoon meerdere kenmerken aanwezig zijn en ook of die persoon wel of niet een bepaalde uitkomst heeft gehad (supervised machine learning). Vervolgens leert de computer op basis van de aangeboden kenmerken te voorspellen wie **welke uitkomst** heeft gehad. Vooral in gevallen waar de uitkomst minder duidelijk te **definiëren** is, neemt de kans op vertekening toe. Neem bijvoorbeeld een uitkomst als kwaliteit van leven, die zich niet makkelijk in een getal laat uitdrukken. Kwaliteit van leven heeft namelijk naast meetbare en daarmee objectiveerbare kenmerken, zoals beperkingen als gevolg van ziekte, veel subjectieve en minder goed te meten kenmerken omvat, zoals ervaren gezondheid, regie over eigen leven, of geluksgevoel. Je kunt dit aan een persoon vragen maar als vragen niet mogelijk is, zoals vaak het geval is in kritische situaties op de IC, dan zou je deze kenmerken kunnen afleiden uit de informatie die wel meetbaar is (categorieën, numerieke waarden), anders kan de computer er niet mee rekenen. Dus stel: je beschikt over data met daarin informatie over arbeidsverleden en boodschappen bij een supermarkt per patiënt. Dan zou je als ontwikkelaar van een voorspellend AI-model ervoor kunnen kiezen om 'werkloosheid' als maat voor regie over eigen leven kunnen nemen en 'consumptie van gezonde producten' als maat voor ervaren gezondheid. Hoewel deze keuzes op het eerste gezicht redelijk lijken, zijn het maar een zeer beperkt aantal kenmerken die iets zeggen over iemands kwaliteit van leven. Bovendien kunnen dergelijke keuzen leiden tot **(onbedoelde) discriminatie**. Zo zullen patiënten met een laag inkomen waarschijnlijk minder vaak gezonde producten kopen, omdat deze producten vaak duurder zijn. Een voorspellend AI-model dat op basis van consumptie van gezonde producten leert, zou dus patiënten met laag inkomen in het hokje "lage kwaliteit van leven" plaatsen en daarmee onbedoeld discrimineren. Vergelijkbare onbedoelde discriminatie zou kunnen voortkomen als een voorspellend AI-model leert dat iemand die werkloos is, minder regie over eigen leven heeft. Zo zouden vrouwen die ervoor kiezen om voor kinderen te zorgen en niet te werken, in het hokje "lage kwaliteit van leven" geplaatst kunnen worden. Als op deze manier vastgestelde kwaliteit van leven vervolgens gebruikt zou worden om kansen op overleving in te schatten, dan kan dat tot zeer **ongewenste classificaties** leiden.

Als het gaat om de **keuze van trainingsdata** dan gaat het met name om de kwaliteit van de historische gegevens en de mate van vertekening die in die gegevens zit (zie hoofdstuk 2). Als er in het verleden suboptimale beslissingen zijn genomen en deze beslissingen zijn in de trainingsdata verankerd, dan zal een voorspellend AI-model deze suboptimale beslissingen reproduceren. Ook is het van belang dat alle bevolkingsgroepen voor welke het model zal worden toegepast in de praktijk, voldoende vertegenwoordigd zijn in de gegevens van een trainingset. Is dat niet het geval dan neemt de kans op vertekening toe voor de **groepen die in de gegevens ondervertegenwoordigd zijn**. Bij de **selectie van gegevens** op basis waarvan uitkomst zal worden voorspeld, kan vertekening plaats vinden zoals hierboven al geïllustreerd. De geselecteerde kenmerken moeten namelijk meetbaar zijn. Dit kan ertoe leiden dat complexe kenmerken vereenvoudigd worden tot één meetbaar kenmerk. Zeker daar waar kosten een rol spelen, zou een ontwikkelaar eerder kunnen kiezen voor kenmerken die makkelijk en snel te achterhalen zijn dan voor kenmerken waarvoor aanvullende dataverzameling nodig is. Stel je hebt inzicht nodig in de gezondheidsvaardigheden van je patiënten, omdat het bekend is dat hoe meer gezondheidsvaardig iemand is des te beter iemand voor zichzelf kan zorgen. Om gezondheidsvaardigheden te kunnen meten zou je een vragenlijst kunnen uitzetten onder de patiënten. Een tijdrovende klus. Het zou kunnen dat een ontwikkelaar dan eerder kiest om in plaats van gezondheidsvaardigheden te meten, een andere karakteristiek gaat gebruiken die makkelijk beschikbaar is zoals het opleidingsniveau van de patiënten. Dit vanuit een aanname dat hoe hoger opgeleid iemand is, hoe hoger zijn/haar gezondheidsvaardigheden. Dit zou een vertekend beeld kunnen geven, omdat bijvoorbeeld het niveau van een opleiding genoten in het buitenland niet goed te vertalen is naar Nederlands opleidingsniveau. Dus voor groepen migranten zou zo'n voorspellend AI-model tot onjuiste conclusies kunnen komen.

Bij proxy 's gaat het om **kenmerken die met elkaar corresponderen** en daarom als elkaars voorspellers kunnen dienen. Deze kenmerken worden vaak gebruikt door voorspellende AI-modellen om mensen in groepen in te delen. Zo heeft frequentie van familiebezoek een sterk verband met het overlijden van een patiënt op de IC. Een voorspellend AI-model zou dus kunnen leren dat hoe hoger de frequentie van IC bezoeken des te hoger de kans op overlijden. Dit terwijl frequentie van bezoeken beïnvloed kan worden door allerlei factoren, zoals mate van bezorgdheid



van de familie, toegang tot vervoer, type werk, eigen gezondheidstoestand. Op deze manier zou een voorspellend AI-model **bepaalde groepen patiënten onbedoeld kunnen benadelen of bevoordelen** en daarmee ongelijkheden creëren of versterken. In tegenstelling tot de keuzes met betrekking tot de uitkomst, trainingsdata en selectie van kenmerken, waar vaak sprake is van menselijke beslissingen, is het veel lastiger om bias op basis van proxy's vast te stellen. Zeker daar waar AI-modellen steeds complexer worden, zoals in geval van neurale netwerken (deep learning).

Bij het ontwikkelen van voorspellende AI-modellen kunnen de boven beschreven mechanismen onbedoeld teweeggebracht worden, maar ze zouden ook bedoeld **misbruikt** kunnen worden om bepaalde groepen te benadelen of te bevoordelen. De AI-ontwikkelaars zouden dus op die manier kunnen maskeren wat hun intenties zijn. Het is van belang dat opdrachtgevers en eindgebruikers van voorspellende AI-modellen **bewust zijn van deze mechanismen** en weten welke keuzes ontwikkelaars van voorspellende AI-modellen hebben gemaakt. De ontwikkelaars dienen hierover transparant te zijn.

### **3.3. Eindverantwoordelijkheid ten aanzien van voorspellende AI-modellen is geborgd.**

Tijdens de interviews werden hoofdzakelijk drie soorten eindverantwoordelijkheden benoemd: de eindverantwoordelijkheid van de hoofdbehandelaar, van de raad van bestuur van een ziekenhuis en van de ontwikkelaars van voorspellende AI-modellen.

Op dit moment is **de hoofdbehandelaar** van een patiënt **eindverantwoordelijk voor medische beslissingen**. Alle geïnterviewden gaven aan dat ze voorspellende AI-modellen als een toevoeging en een hulpmiddel zien en niet als vervanging van de arts. Ze verwachten ook geen feilloos systeem, maar een systeem dat gemiddeld genomen samen met de arts tot betere beslissingen leidt in vergelijking met beslissingen van artsen alleen. Dus ook in het geval dat een voorspellend AI-model gebruikt zou worden ter ondersteuning van medische beslissingen, vindt men dat de hoofdbehandelaar eindverantwoordelijk zou zijn. Van de hoofdbehandelaar wordt wel verwacht dat hij of zij voldoende kennis heeft van de statistiek achter een voorspellend AI-model en zich vergewist heeft dat het gebruik van een voorspellend AI-model veilig is. Het veronderstellen van kennis over de statistiek achter een voorspellend AI-model wordt problematischer naarmate de methode die gebruikt wordt complexer en ondoorzichtiger wordt. Er zijn diverse soorten machine learning methoden beschikbaar waarmee een computer voorspellende modellen kan maken. Een classificatieboom (soms een beslisboom genoemd) is een simpel en voor de mens begrijpelijk model. Meer **complexe varianten** van machine learning modellen zijn neurale netwerken, met name netwerken met veel lagen (deep learning). Neurale netwerken worden daarom ook black-box modellen genoemd, omdat het voor de mens zeer lastig te achterhalen en doorgronden is waarop de computer zijn berekeningen baseert en hoe de computer tot een resultaat komt.

Het is de verantwoordelijkheid van **een raad van bestuur van een ziekenhuis** om op een wetenschappelijke en maatschappelijk verantwoorde wijze voorspellende AI-modellen in gebruik te nemen en te blijven monitoren of het gebruik veilig is. Daarnaast dient een bestuur te zorgen dat hun personeel voldoende kennis en vaardigheden heeft om met de nieuwe technologieën, zoals AI, verantwoord om te kunnen gaan. Zo houdt de Inspectie Gezondheidszorg en Jeugd (IGJ) toezicht op veilig gebruik van medische technologie die gebruikt wordt bij de diagnose, behandeling en ondersteuning van ziekten en gebreken. Veilige toepassing van medische technologie betekent een veilig product, in handen van een getrainde gebruiker, in een omgeving die veilig gebruik kan garanderen. Het convenant '**Veilige toepassing van medische technologie in het ziekenhuis**' geeft invulling aan de risicobeheersing en de veilige toepassing van medische technologie binnen de ziekenhuiszorg. De raad van bestuur is verantwoordelijk voor de organisatie en de uitvoering van de in dit convenant genoemde procedures. Daarnaast dient de raad van bestuur ervoor te waken dat het gebruik van patiëntengegevens conform geldende regelgeving gebeurt. Recent kwam het Engelse Royal Free Trust in London (onderdeel van National Health System) in opspraak nadat New Scientist een contract met Google onder ogen kreeg, waarin 5 jaar aan patiëntengegevens in bulk (enkele miljoenen) met Google zouden worden gedeeld om onder andere voorspellende AI-modellen te ontwikkelen. Dit bleek niet conform AVG.

Voor meer informatie over geheimhouding en privacy in context van big data, verwijzen we graag naar het rapport van Neuteboom & Sijmons.

**De verantwoordelijkheid van de ontwikkelaars van voorspellende AI-modellen** is om zorgvuldig met de gegevens om te gaan, de eindgebruikers zo goed mogelijk voor te lichten over hoe een voorspellend AI-model is ontwikkeld en de daaruit voortkomende beperkingen van een voorspellend AI-model kenbaar te maken. Een ontwikkelaar kan een zogenaamde disclaimer formuleren waarin deze informatie terug te vinden is. In tegenstelling tot juridische en professionele verantwoordelijkheid van medici die verankerd is in de wetgeving, is **de verantwoordelijkheid van de ontwikkelaars** van voorspellende AI-modellen, op AVG na, op dit moment **niet op een vergelijkbare wijze gegrondvest**. Ook hier geldt dus dat er op bestuurlijk en project niveau een duidelijk eisenpakket moet komen ten aanzien van de verantwoordelijkheid van ontwikkelaars van voorspellende AI-modellen.

### **3.4. Passende vergoeding naar rato van de investering in de ontwikkeling van voorspellende AI-modellen is contractueel vastgesteld.**

Zoals eerder genoemd, dient een computer voorspellende AI-modellen te leren op basis van gegevens van voldoende datakwaliteit. Dit betekent dat de te gebruiken gegevens eerst voorbereid moeten worden, in lijn met de doelen waarvoor een model ontwikkeld gaat worden. **Vorbereiden van gegevens**, ook wel pre-processen genoemd, houdt onder andere in: het opschonen, valideren, transformeren, samplen en labelen van gegevens. Deze voorbereiding neemt meestal **meer dan de helft van de ontwikkeltijd in beslag** en vraagt om input van experts met inhoudelijke kennis. In de zorg gaat het dan meestal om experts met **medische kennis**, zoals artsen. Bij een samenwerking met vooral commerciële partijen, is het daarom verstandig om een passende vergoeding af te spreken voor de inzet van artsen bij het voorbereiden van de gegevens. Deze vergoeding moet in verhouding staan tot de prijs die een commerciële partij in de toekomst wil vragen voor het ontworpen model. In de interviews is enkele keren de farmaceutische industrie als een voorbeeld genoemd van hoe het mis kan gaan. Bij de ontwikkeling van nieuwe geneesmiddelen maakt de farmaceutische industrie stevig gebruik van publieke middelen zoals: ziekenhuisfaciliteiten, ziekenhuismiddelen en expertise van artsen en verpleegkundigen. Echter, de vergoeding die de publieke sector krijgt voor deze diensten staat in schril contrast tot de vaak exorbitante prijzen die de farmaceutische industrie vraagt voor haar met publieke middelen ontwikkelde nieuwe geneesmiddelen.

### **3.5. Intellectueel eigendom ten aanzien van voorspellende AI-modellen is geborgd.**

In de lijn met de afspraken over vergoeding voor de ontwikkeling van een model, is het van belang om ook duidelijke afspraken te maken over **intellectueel eigendom**. Het eigendomsrecht biedt een uitvinder of ontwikkelaar van een voorspellend AI-model bescherming tegen onrechtmatige acties van derden: niemand mag dit model zonder toestemming gebruiken. Dit houdt ook vaak in dat de eindgebruikers geen inzicht krijgen in hoe een model precies werkt, zodat ze deze niet na kunnen maken. Vooral **commerciële partijen** trachten op deze wijze winsten te maken met voorspellende AI-modellen. Een bekend voorbeeld is op AI-gebaseerde Watson for Oncology van IBM: een systeem dat in staat is om vrije tekst te lezen en analyseren uit o.a. medische boeken en wetenschappelijke artikelen, om artsen te ondersteunen bij het vinden van evidence-based behandelingen in de oncologie. Een belangrijke kanttekening hierbij is dat Watson for Oncology niet uit zichzelf op basis van de informatie uit boeken en artikelen tot adviezen kan komen. De adviezen van Watson for Oncology zijn gebaseerd op vele uren training en invoer van medische gegevens door artsen van Memorial Sloan-Kettering Cancer Center in de Verenigde Staten. Dit betekent dat de meerwaarde van Watson for Oncology voor andere ziekenhuizen en andere landen beperkt is omdat o.a. richtlijnen en/of zorgsystemen anders zijn. Dit betekent dus ook dat als je Watson for Oncology aanschaft als ziekenhuis, dat er van je artsen om een **behoorlijke tijdsinvestering wordt gevraagd om het systeem te trainen** en ook om een investering voor data pre-processing. De getrainde Watson wordt waarschijnlijk weer door IBM doorverkocht met kennis vanuit de organisatie. Hier ziet de organisatie zelf qua winst op basis van het geleverd intellectueel eigendom waarschijnlijk niks van terug.

## 4. RANDVOORWAARDEN TEN AANZIEN VAN ONTWIKKELING EN EVALUATIE VAN VOORSPELLENDE AI-MODELLEN

*“Ik zou zeker een stap-voor-stap overzicht willen hebben van hoe een model tot zijn voorspelling komt en op zijn minst bijvoorbeeld een overzicht van variabelen die als de sterkste voorspellers zijn geoormerkt. Als ik totaal niet meer kan herleiden waarop een berekening is gebaseerd is, dan zou ik het niet gebruiken. Ik voel me daar niet prettig bij. Naast zo een overzicht wil ik ook weten wat de beperkingen zijn van het model, dus bijv. welke patiëntengroepen zijn onvoldoende vertegenwoordigd in de gegevens waarop een model zijn berekening baseert.”*

*“Je moet nagaan hoeveel zekerheid mensen willen hebben om daadwerkelijk consequentie te verbinden aan de prognose dat zo een model berekend. Dat gaat namelijk nooit 100% zekerheid zijn, maar 80% is misschien wel te weinig.”*

*“Als AI leidt tot in-transparantie dan hebben we ‘doctor knows best’ vervangen door ‘machine knows best’. Dat lijkt mij afschuwelijk.”*

*“Ik hoop dat voorspellende AI-modellen worden ontwikkeld door wetenschappers in samenwerking met artsen. En dat daar zorgvuldig over nagedacht wordt, gevalideerd en geëvalueerd.”*

*“Als een ziekenhuis zo een model ontwikkeld heeft, dan dient zo een model open source/open access te zijn. Zeker als het met wetenschapsgeld ontwikkeld is.”*

*“We moeten er ook voor open staan dat we niet alles kunnen begrijpen, want dan heb je AI niet nodig. Maar ik zou in ieder geval wel willen weten welke aannames zo een model doet over bepaalde patiënt categorieën, leeftijd, diagnoses etc. Die grote aannames, die moet je wel kunnen snappen.”*

*“En er moet een feedbackmechanisme in zitten om terug te geven wat er daadwerkelijk is gebeurd en ook hoe de patiënt bijv. daadwerkelijk zijn/haar kwaliteit van leven ervaart na het ontslag uit het ziekenhuis.”*



#### 4.1. Voorspellende AI-modellen zijn ontwikkeld en geëvalueerd conform wetenschappelijke richtlijnen.

In het artikel van Wyatt et al. *Prognostics models clinically useful or quickly forgotten?* wordt stapsgewijs uiteengezet waarom artsen de tot nu toe gepubliceerde voorspellende modellen in de dagelijkse praktijk niet tot nauwelijks gebruiken. De aangevoerde hoofdredenen zijn:

- Gebrek aan klinische geloofwaardigheid.
- Gebrek aan bewijs dat een voorspellend model medische beslissingen kan ondersteunen, dat wil zeggen bewijs van accuraatheid, en ten aanzien van generaliseerbaarheid en vooral effectiviteit.

Een voorspellend model wint aan klinische geloofwaardigheid indien: 1) alle **klinisch relevante gegevens** getoetst zijn voor toevoeging als kenmerk aan een model, 2) de benodigde gegevens **makkelijk en betrouwbaar te verzamelen** en te registreren zijn voor artsen, om zodoende tijdig medische beslissingen te kunnen ondersteunen; 3) **arbitraire afkapwaarden** voor continue variabelen zijn **vermeden**, 4) de wijze waarop een model tot zijn berekening komt en een voorspellend model produceert **begrijpelijk en logisch zijn voor de artsen**, die er op moeten vertrouwen, 4) de voorspelling **makkelijk uit te leggen valt aan een patiënt**.

Als het gaat om bewijs van **accuraatheid**, dan zal een voorspellend model niet snel van meerwaarde worden gezien als zijn voorspellingen niet ten minste even accuraat zijn als die van de artsen die het zouden gebruiken. Een voorspellend model moet zelden falen bij het voorspellen van een uitkomst die plaats zal vinden (lage vals negatieve waarde) maar ook zelden falen om een voorspelling te maken van een uitkomst die niet plaats zal vinden (lage vals positieve waarde). Zeker in het geval van medisch beslissingen aangaande leven of dood, is een sterk onderscheidend vermogen van een model (het kunnen onderscheiden van patiënten op basis van een bepaalde uitkomst) een must. Oftewel, de vals positieve waarde (zeggen dat een patiënt een bepaalde uitkomst heeft, terwijl dit niet waar is) zou bij voorkeur nul moeten zijn, met een zeer smal betrouwbaarheidsinterval.

Deze waarden moeten worden gecheckt op **statistisch voldoende hoeveelheid aan gegevens**, waarin aanwezigheid of afwezigheid van de uitkomst op een betrouwbare wijze is vastgesteld. Als een model een kans op een uitkomst voorspelt, bijv. 40% kans op overleving van IC opname, dan moet die kans ook accuraat zijn (oftewel 40% van patiënten met deze kans overleeft daadwerkelijk de IC opname). Er zijn maar weinig voorspellende modellen bekend in de wetenschappelijke literatuur die op deze manier zijn geëvalueerd.

Een voorspellend model dient ook breed toepasbaar, of met andere woorden **generaliseerbaar** te zijn. Dat wil zeggen: Een voorspellend model is rigoureuus getest op een nieuwe set van patiëntengegevens, nieuw in termen van tijd (bijv. uit een ander jaar) en plaats (bijv. uit een ander ziekenhuis). Nieuw betekent ook dat het bewijs van accuraatheid niet alleen op basis van trainingset is vastgesteld, maar ook op basis van een testset met gegevens die een voorspellend model nog nooit "heeft gezien". Een analogie: Stel je krijgt een oefententamen met de juiste antwoorden om thuis te oefenen en op het echte tentamen krijg je weer hetzelfde oefententamen. De kans is groot dat je een beter cijfer krijgt voor dit tentamen, dan op een ander niet eerder gezien tentamen. Het doel van een tentamen is te testen of je kunt generaliseren, dus toepassen van je kennis op nieuwe situaties. Hetzelfde geldt voor het leren van voorspellende AI-modellen. Het testen op ongeziene data heet validatie.

Daarnaast dienen alle gegevens die een voorspellend AI-model gebruikt duidelijk gedefinieerd te zijn conform wijd geaccepteerde definities. Ook dient een voorspellend model bij voorkeur conform een protocol prospectief ontwikkeld en geëvalueerd te worden, op een groep patiënten die geselecteerd is op basis van kenmerken van patiënten waarvoor een model vaak toegepast zou worden in de toekomst. Een treffende vergelijking wordt gemaakt met onderzoek naar effectiviteit van geneesmiddelen, waarbij wordt gesteld dat geen arts een geneesmiddel aan een patiënt zou voorschrijven als zo'n geneesmiddel alleen in een reageerbuis was onderzocht en nooit bij mensen via gecontroleerde geblindeerde klinische trials (gouden standaard). Als het gaat om voorspellende AI-modellen, zouden zorgverleners even **gedegen wetenschappelijke**

**bewijsvoering** voor hun effectiviteit moeten eisen. Bij voorkeur via gecontroleerde gerandomiseerde klinische trials, waarin de effecten van een voorspelling op klinische processen en/of patiënt uitkomsten worden onderzocht in vergelijking met een situatie zonder een voorspelling. Of in woorden van [Isaac Kohane](#), hoogleraar biomedische informatica van Harvard Medical School: “*Computing may be the big 21st-century drug – let’s treat it just as responsibly.*”

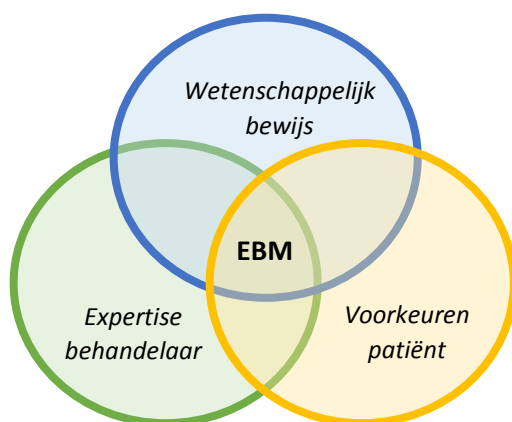
Om (wetenschappelijke) publicaties over voorspellende modellen in het algemeen en voorspellende AI-modellen in het bijzonder **kritisch te kunnen beoordelen**, kan gebruik gemaakt worden van meerdere richtlijnen waaronder:

- <https://www.strobe-statement.org>
- <https://www.record-statement.org>
- <https://www.tripod-statement.org>
- [Lou et al. Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research: A Multidisciplinary View. J Med Internet Res 2016; 18\(12\): e323](#)
- [Riley et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. BMJ. 2016 Jun 22;353:i3140](#)

In het artikel van Shah et al. over big data en predictive analytics in de gezondheidszorg wordt zelfs gesuggereerd om een **onafhankelijke instantie** op te zetten om voorspellende AI-modellen en hun toepassing in de klinische praktijk te **certificeren**. Daarmee kan geborgd worden dat dergelijke modellen hun beloften waarmaken wat betreft de toegevoegde waarde en de uitkomsten voor de individuele patiënt en het gehele gezondheidszorgsysteem. Immers, de potentie van voorspellende AI-modellen om medische beslissingen te beïnvloeden impliceert ook de potentie om schade te veroorzaken door onjuiste informatie op het moment van besluitvorming. Hoewel een slecht voorbereide arts ook fouten kan maken bij het nemen van medische beslissingen voor één patiënt per keer, kan een slecht gevalideerde en geëvalueerde voorspellend AI-model veel meer patiënten tegelijkertijd schade toebrengen.

#### 4.2. De uitkomsten van voorspellende AI-modellen zijn uitlegbaar.

De geïnterviewden dachten verschillend over de mate waarin de berekeningen achter een voorspellend AI-model te begrijpen zouden moeten zijn door de eindgebruikers. Ondanks dat, was er echter ruime consensus over het feit dat een arts op zijn minst op hoofdlijnen zou moeten kunnen zien en begrijpen op basis van welke (set van) gegevens/aannames een voorspellend AI-model de uitkomst heeft gebaseerd. Deze eis raakt in essentie aan de principes van **Evidence Based Medicine (EBM)**. EBM is een integratie van drie bronnen van informatie bij het nemen van medische beslissingen: 1) systematische synthese van wetenschappelijk bewijs (evidence), 2) de expertise van de behandelaar en 3) de voorkeuren van de patiënt (zie Figuur 1).



**Figuur 1:** Schematische weergave van Evidence Based Medicine (EBM).

Om EBM-principes in stand te kunnen houden moet de behandelaar in staat worden gesteld om het bewijs te interpreteren, en te beoordelen in hoeverre het van toepassing is op een situatie van de individuele patiënt die hij of zijn behandeld. Vervolgens moet een behandelaar zijn interpretatie van het bewijs, aangevuld met eigen expertise, uit kunnen leggen aan de patiënt. Hierdoor wordt de patiënt in staat gesteld om eigen voorkeuren mee te laten wegen bij het nemen van een beslissing. Op het moment dat een behandelaar niet meer in staat is om de uitkomst van een voorspellend AI-model zelf te begrijpen, te interpreteren en aan de patiënt of naasten uit te leggen, komen de EBM-principes in gevaar.

#### **4.3. Voorspellende AI-modellen zijn bij voorkeur multicenter ontwikkeld en geëvalueerd**

Deze eis komt voort uit de behoefte van artsen dat een voorspellend AI-model **breed toepasbaar** moet zijn (zie paragraaf 4.3). Hiernaast kan dit zorgen dat mogelijke vertekeningen zichtbaar worden, zodat ervan geleerd kan worden tussen de ziekenhuizen (zie Paragraaf 2.2 & 2.5).

#### **4.4. Voorspellende AI-modellen worden als open source modellen beschikbaar gesteld.**

Open source betekent in het kort: vrije toegang geven tot de broncode. In geval van voorspellende AI-modellen gaat het over de rekenregels achter het model. Een model dat via open source beschikbaar is gemaakt, stelt de (potentiële) eindgebruikers in staat om na te kunnen gaan welke statistiek achter een voorspellend AI-model schuilgaat, en dus hoe de berekeningen tot stand komen. Echter, voor voorspellende AI-modellen die via machine learning ontwikkeld worden, is dit vaak niet eenvoudig te realiseren. Dit komt doordat dergelijke modellen niet makkelijk interpreteerbaar zijn (denk aan een neurale netwerk met vele lagen), en sommige modellen kunnen voortdurend leren op basis van nieuwe data die ze aangeboden krijgen. Om up-to-date rekenregels te kunnen vrijgeven, zouden daarom ook patiëntgegevens op basis waarvan het model leert, vrij toegankelijk gemaakt moeten worden. Een onwenselijke situatie. Wat wel mogelijk is, is het openstellen van de rekenregels achter de finale versie van een voorspellend AI-model, zoals bekend ten tijde van eindevaluatie.

Daarnaast zouden voorspellende AI-modellen ter ondersteuning van medische beslissingen **gratis** beschikbaar moeten zijn. Indien ontwikkeld met grotendeels publieke middelen ligt gratis beschikbaar stellen van broncodes en AI-modellen voor de hand. Dit gebeurt meestal via wetenschappelijke publicaties. Voor AI-modellen die ontwikkeld zijn door commerciële partijen ligt dit anders.

## 5. RANDVOORWAARDEN TEN AANZIEN VAN GEBRUIK VOORSPELLENDE AI-MODELLEN IN DE KLINISCHE PRAKTIJK

*"We zijn natuurlijk gebaat bij snelle en accurate prognose zodat de familie ook weet waar ze aan toe zijn. Maar dat leidt niet altijd tot een medische beslissing op dat moment. En vaak moet zo een besluit ook bij de dokters bezinken"*

*"Ik denk niet dat we het schrikbeeld gaan halen van een computer die beslissingen neemt, maar ik denk wel dat zo een model steeds zwaarder gaat wegen dan het gevoel van de arts."*

*"Je ziet dat mensen vaak causale conclusies trekken uit dit soort voorspellende modellen: zo van als ik dan één variabele verander, dan komt er iets anders uit. Dat kan niet. Dat grensvlak is heel dun: alleen predictie en kans hebben, versus meteen behandeling afbouwen. Dat klopt niet. Dan gaat het hopeloos mis, dan is AI niet geschikt om in te zetten. Dus het is niet zozeer de techniek, maar wat mensen er mee doen. En hoe die techniek gaat worden ingezet bij een bepaalde beslissing."*

*"Als arts heb je soms 'last van' dat je je altijd bepaalde patiënten herinnert. Een heel vervelend verhaal van een vader met jonge kinderen, of die ene patiënt waarvan je denkt die gaat het nooit redden, die het toch heeft gered. Of een patiënt waarvan je alles op alles hebt gezet en achteraf denkt: dit hadden we anders moeten doen, die patiënt heeft nog zoveel geleden aan het eind van z'n leven. Die verhalen verstoren je herinnering. Daar heeft de computer geen last van."*

*"Zo een model zie ik als een aanvulling op de expertise van de arts. Expertise van de arts daar kan een computer niet tegenop."*

*"Ik vraag me af of de artsen allemaal wel voldoende kennis hebben om die modellen zelf te interpreteren."*

*"Wat bij artsen vaak meespeelt is de zogenaamde 'escalation commitment': als je een band hebt met de familie of je bent een chirurg die al drie keer heeft geopereerd, dan wil je langer door behandelen. Zo een model geeft dan aan: deze patiënt is 90, hij heeft drie operaties gehad, zijn nieren stoppen nu, het is niet zinvol om door te behandelen. Maar die chirurg zal misschien denken: ik heb net nog geopereerd, kom we proberen het nog even. Dus ja, AI is in dat opzicht wel neutraler."*

*"Een model maakt een voorspelling van de kans op overleving en deze voorspelling zien we dan terug in het EPD, net als dat we nu naar de laboratoriumuitslagen kijken. Gewoon, één van de vele andere aanwijzingen die we gebruiken om te beslissen wat te doen. Een extra bron van informatie dat helpt een waarde te geven aan wat je voelt."*

*"Een model dat elke dag een soort real-time meelopend grafiekje met prognose maakt en dat in het EPD toont, zodat je als arts kunt zien nu is de kans op overleving van deze patiënt zoveel procent. En als elke dag beetje beter wordt, gaan we de goede kant op. Op die manier ga je ook niet vastpinnen op één getal, maar dat je de trend gaat zien."*

## 5.1. Voorspellende AI-modellen dienen enkel als beslissingsondersteuning.

De uitkomsten van voorspellende AI-modellen dienen enkel gebruikt te worden als **hulpmiddel om medische beslissingen te ondersteunen**. Een menselijke eindbeslissing moet altijd leidend blijven. Zeker in het geval van complexe beslissingen, zoals wel of niet een levensbehoudende behandeling op de IC staken. Een goede beslissingsondersteuning is een ondersteuning waarbij informatie door een voorspellend AI-model, aangevuld met informatie van de arts en patiënt of diens naasten, tot betere beslissingen leidt in de dagelijkse praktijk. Een voorbeeld van zo'n **integratie van expertise van zorgverleners en een voorspellend model** is de SUPPORT-studie: Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments (SUPPORT) van Knaus et al. In deze studie werd een voorspellend model ontwikkeld en geëvalueerd waarmee een schatting van overleving na 180 dagen werd berekend voor ernstig zieke ziekenhuispatiënten. De schattingen die gemaakt werden door de artsen en door het model (afzonderlijk van elkaar), bleken vrijwel identiek. Echter, wanneer de schattingen van de artsen werden geïntegreerd in het SUPPORT-model, als een additionele variabele, verbeterde de voorspellende accuraatheid van het model significant evenals het vermogen om patiënten te identificeren met hoge waarschijnlijkheid op overleving of overlijden. Met andere woorden: hoewel de huidige discussie over toepassing van AI vooral in het teken staat van een "mens versus computer" framing, is de vraag "hoe kan interactie tussen mens en computer zinvol worden vormgegeven?" veel relevanter. Immers de impact van een voorspellend AI-model hangt niet alleen af van de eigenschappen van een model, maar ook van hoe zo'n model door de mensen wordt gebruikt.

Desalniettemin zijn er ook succesvolle toepassingen van **AI-modellen beschikbaar die onafhankelijk van de arts besluiten nemen**. Tijdens de interviews kwam kunstmatige beademing op de IC ter sprake. De nieuwste generatie kunstmatige beademingsmachines wordt aangestuurd door AI-modellen, waarbij de beademingsmachine zich op basis van diverse metingen continu aanpast aan de (veranderende) behoefte van de patiënt. Dit zonder tussenkomst van de arts. Dit voorbeeld laat zien dat als het gaat om ondersteuning van gezondheidszorg door AI-modellen, de mate van autonomie van AI-systemen mogelijk per toepassing zou moeten worden afgestemd met de beoogde eindgebruikers.

Een dergelijke aanbeveling werd gedaan door Proud et al., onderzoekers van NASA in de Verenigde Staten. Deze onderzoekers hebben een beoordelingsschaal voor het niveau van autonomie ontwikkeld, the **Level of Autonomy (LOA) Assessment Tool**. The LOA Assessment Tool heeft als doel om het meest optimale niveau van autonomie vast te stellen voor een specifieke toepassing/functie binnen een Autonomous Flight Management system. De niveaus van autonomie variëren van niveau 1 tot niveau 8 en corresponderen respectievelijk met complete menselijke tot complete computer verantwoordelijkheid. De onderzoekers hebben elk niveau van autonomie afgestemd op functietypes Observeren, Oriënteren, Beslissen, Handelen. Volgens LOA Assessment Tool valt de hierboven beschreven kunstmatige beademingsmachine in niveau 6 van functietype Handelen: *"de computer voert handelingen automatisch uit; informeert de mens; de mens kan computerhandelingen na uitvoering overrulen; de mens monitort op de achtergrond op onvoorziene omstandigheden."* Een voorspellend model voor de kansen op overleving van IC-opname past volgens de wensen van de geïnterviewden bij een veel lager niveau, namelijk niveau 3 van functietype Oriënteren: *"een computer analyseert de gegevens en maakt voorspellingen; de mens is verantwoordelijk voor de interpretatie van de voorspelling."*

## 5.2. Uitkomsten voorspellende AI-modellen zijn direct beschikbaar in het EPD.

Een sterke voorkeur is uitgesproken voor de **integratie van voorspellende AI-modellen in het EPD**. De berekeningen moeten gebaseerd zijn op up-to-date informatie, zodat de berekende voorspelling meegenomen kan worden tijdens dagelijkse beslismomenten op de afdeling en/of contactmomenten met patiënten en/of naasten. Als het gaat om continu leren van een voorspellend AI-model dan zouden er mogelijkheden beschikbaar moeten zijn voor eindgebruikers om aan te geven of de voorspelling van een model klopte of niet, en dat deze feedback meegenomen wordt om de voorspellingen continu te optimaliseren voor de toekomstige patiënten. Daarnaast, en mogelijk specifiek voor een voorspellend AI-model die de kans op overleving van IC-opname zou gaan berekenen, gaven de artsen en patiënten de voorkeur aan een voorspellend



AI-model die een **trend** zou kunnen tonen over tijd. Een trend zegt in dit geval meer dan één moment opname. Bovendien, zeker in geval van dergelijke complexe beslissingen, worden er zelden op basis van één meting beslissingen genomen. Meestal wordt de patiënt gedurende meerdere dagen geobserveerd en op basis daarvan worden vervolgstappen bepaald.

### **5.3. Eindgebruikers beschikken over voldoende kennis en vaardigen om voorspellende AI-modellen verantwoord toe te kunnen passen.**

Zoals vermeld in paragraaf 3.5 worden er eisen gesteld aan eindgebruikers van medische technologie in de zorg. Deze eisen staan beschreven in het convenant **“Veilige toepassing van medische technologie in het ziekenhuis.”** Deze eisen gelden onder andere voor beslissingsondersteunende systemen en daarmee ook voor voorspellende AI-modellen die bedoeld zijn om medische beslissingen te ondersteunen. Tijdens de interviews waren het vooral de technologie medische experts, maar ook patiënten, die zich afvroegen in hoeverre artsen in staat zijn om verantwoord om te kunnen gaan met AI-technologie. Medisch ethicus prof. dr. Hans van Delden was uitgesproken huiverig over een situatie waarin artsen de uitkomsten van voorspellende AI-modellen als de waarheid gaan zien en zich er afhankelijk van maken. Het is volgens hem daarom noodzakelijk dat een arts kan beoordelen of een voorspellend AI-model kwalitatief goed is ontwikkeld en welke beperkingen en onzekerheden erin zitten. Alleen met deze kennis kan een arts vervolgens nagaan of de kans, zoals getoond door een voorspellend AI-model, past bij het individu met z'n unieke omstandigheden. Dit om vervolgens zijn/haar interpretatie in gesprek met de patiënt te verwoorden en na te gaan wat de wensen zijn van de patiënt.

In het artikel van dr. Marjolein Geurts et al. over end-of-life beslissingen bij patiënten met ernstige acute hersenschade, spreekt men zijn verontrusting uit over de beperkte aandacht die besteed wordt aan het significante risico op schade door toepassing van voorspellende modellen. Schade in termen van **verspilling van middelen** en **schade aan patiënten**, doordat onnauwkeurige voorspellingen tot slechtere beslissingen kunnen leiden. De auteurs stellen dat het onwaarschijnlijk is dat klinici en ziekenhuisbestuurders **voldoende methodologische kennis** in huis hebben om voorspellende AI-modellen (die ze kopen) op hun meerwaarde en risico's kritisch te kunnen beoordelen. Zo trok M.D. Anderson Cancer Center in Houston, Texas na 5 jaar en 62 miljoen dollar verder, de stekker uit Watson for Oncology project, omdat het niet voldeed aan de verwachtingen, en de beloftes van IBM niet waar werden gemaakt. Een pijnlijke les voor ziekenhuisbestuurders en een reden om expertise in huis te hebben, om eisen aan dergelijke systemen van tevoren goed te definiëren.

Een **algemene training voor zorgverleners en bestuurders** over kansen en beperkingen ten aanzien van machine learning en hergebruik van routinematig verzamelde gegevens, zou mogelijk kennishiaten bij deze eindgebruikers kunnen verminderen en meer bewustzijn doen creëren. Een voorbeeld van een dergelijke training is een e-learning module [Health Data Science/Big data](#), ontwikkeld door de afdeling Klinische Informatiekunde van Amsterdam UMC. In het Themanummer Medisch Specialist 2025 stelt Marjolein Kremers, internist in opleiding en voorzitter van De Jonge Specialist, dat het cruciaal is dat elke arts (in opleiding) leert hoe je de kwaliteit van verschillende technologische innovaties kritisch leert beoordelen. Dit om onderscheid te maken tussen zinvolle en zinloze innovatie en om beter aan te kunnen geven wat nodig is om het werk van de arts goed te kunnen doen. Het betrekken van methodologen, machine learning experts en ethici bij het vaststellen van ziekenhuisbreed beleid ten aanzien van ontwikkeling, evaluatie en implementatie van voorspellende AI-modellen, is tevens geen overbodige luxe.

Voor burgers is er recent een gratis online [Nationale AI-cursus](#) beschikbaar gekomen. Deze cursus is mede mogelijk gemaakt door het Ministerie van Binnenlandse Zaken en Koninkrijksrelaties. Echter, deze cursus is niet specifiek voor de medische sector waardoor aspecten van AI die zorgspecifiek zijn niet aan bod komen.

### **5.4. Voorspellende AI-modellen worden na implementatie in de klinische praktijk continu geëvalueerd**

Een belangrijke voorwaarde voor het verantwoord gebruik van voorspellende AI-modellen in de klinische praktijk is een **continue monitoring op klinische meerwaarde en (potentiële) risico's** die het gebruik van dergelijke voorspellende AI-modellen met zich mee brengen. Het begint

natuurlijk met het in gebruik nemen van kwalitatief goede en wetenschappelijk verantwoord ontwikkelde voorspellende AI-modellen. Echter, na het in gebruik nemen van zo'n model kunnen zich onvoorziene situaties voordoen waarmee niet eerder rekening is gehouden. Je zou kunnen stellen dat voor voorspellende AI-modellen een vergelijkbaar traject in gang gezet zou moeten worden zoals dat gebeurt via post-marketing onderzoek van nieuwe geneesmiddelen. Tevens zou de zogenaamde **plan-do-check-act cyclus** de basis moeten vormen van elke innovatie in de zorg.

Op de vraag of **normen en waarden** in voorspellende AI-modellen ingebouwd zouden moeten worden, antwoorde merendeel van de geïnterviewden "nee". Dit omdat men voorspellende AI-modellen ziet als ondersteuning en niet als beslisser. Het is aan de mensen die met voorspellende AI-modellen werken en aan diegenen waarop de uitkomsten van toepassing zijn, om te toetsten welke normen en waarden van belang zijn op dat moment (qua plaats en tijd) en om dus **continu bij te sturen** waar nodig. Er moet ruimte overblijven voor verandering van de moraal. Tegelijkertijd kan de inzet van **methoden om bias in de data op te sporen** helpen om bijvoorbeeld discriminatie op basis van gender, leeftijd, etniciteit, religie, seksuele oriëntatie of ander eigenschappen door voorspellende AI-modellen tegen te gaan. In het proefschrift van Julius A. Adebayo "*FairML: ToolBox for diagnosing bias in predictive modeling*" wordt een FairML toolbox beschreven die kan worden gebruikt om voorspellende AI-modellen te auditen op discriminatie. Dit wordt bewerkstelligd door de relatieve betekenis van de gegevens die het model gebruikt bij het bepalen van de uitkomst te kwantificeren.

## 6. RANDVOORWAARDEN TEN AANZIEN VAN COMMUNICATIE VAN VOORSPELLINGEN VAN AI-MODELLEN AAN DE PATIËNT

*"Ik vind dat het wel zo eerlijk is als de artsen je vertellen dat ze zo een model hebben gebruikt. Ik denk dat openheid altijd de relatie tussen arts en patiënt verbetert. Is geen zwakte ofzo als een arts zo een model gebruikt."*

*"Als het gesprek dat we nu hebben zou veranderen in een meer mathematisch gesprek, dan komt dat niet te goede aan het gesprek tussen arts en patiënt."*

*"Ik zou wel willen weten op welke IC-afdelingen in Nederland zo een model wordt gebruikt. Een IC met zo een model geeft mij iets meer geruststelling dat er goed wordt nagedacht over behandelkeuzes."*

*"Ik hoef het niet per se aan de familie te vertellen dat ik een model heb gebruikt en wat eruit komt. Dat is niet de meest handige manier. We zeggen nooit tegen de familie van: uw naaste heeft een prognose van 53% om er goed of slecht uit te komen. Dat komt ook niet over. Je zegt ook nooit je hebt nog drie maanden en twee weken te leven. Dat weet je gewoon niet."*

*"Bij zo een belangrijke beslissing zou ik niet genoeg nemen met alleen globaal uitleg over een prognose in percentages. De arts moet het ook begrijpen en uitleggen hoe dat percentage tot stand is gekomen."*

*"Pas als een model heel betrouwbaar voorspelt dan pas kun je het gebruiken richting patiënt en familie. Je moet daar wel heel genuanceerd over praten. En je ook afvragen, of je mensen daarmee niet te veel belast. Hoeveel informatie is ook daadwerkelijk zinvol? Dat is iets wat je per patiënt moet bekijken."*

*Ik vind personalized medicine een misleidend woord. Ik spreek liever van precision medicine, omdat je preciezer kunt voorspellen welke parameters samenhangen met welke uitkomst. Om het personalised te maken heb je de wensen van de patiënt en/of naasten nodig. En dan kun je sturen op een individueel relevante uitkomst."*

*"Ik denk dat het gesprek met naasten over wel of niet doorbehandelen veel duidelijker wordt als je de uitkomst van een model kan laten zien en kunt zeggen: deze en deze factoren worden gewogen en leiden tot deze prognose. Normaal zeg je ook iets over prognose maar dan met veel woorden, en dan zou je het iets concreter kunnen maken."*

*"Ook al heb je straks een model dat heel goed in staat is om kansen op overleving te voorspellen dan nog moet je de familie ook wel ruimte geven om aan die prognose te wennen. Je moet dan wel nadenken over hoeveel ruimte. Dat klinkt een beetje hard, maar ons primaire uitgangspunt is niet zinloos handelen en first do no harm. Dus we moeten niet een patiënt schade berokkenen, omdat het voor de familie fijner is. Dat is een goed gesprek wat je moet hebben."*

*"Ik denk dat het gebruik van zo een model de relatie tussen arts en patiënt zal veranderen. Maar die relatie die verandert altijd door de jaren heen, het wordt steeds volwassener. Het gaat nog steeds wel om vertrouwen hoor. Er is niets zo belangrijk in je leven als je gezondheid. En hoe je heel snel je hele vertrouwen weggeeft en zegt: ik weet het ook niet meer, jij bent hier de arts. Dan is het wel fijn als ik mijn oordeel als arts kan toetsen aan wat zo een model berekent en dan sterker sta richting de patiënt of de naasten."*



## **6.1. De presentatie van voorspellingen van AI-modellen moet aansluiten op het niveau van begrip, gezondheidsvaardigheden en individuele wensen van patiënten en/of hun naasten.**

Bij voorkeur worden de voorspellingen van AI-modellen gebruikt als een **informatiebron tijdens het gesprek met de patiënt of naasten**. Omdat de mate waarin patiënten en naasten in staat zijn om medische informatie te verwerken en te begrijpen zeer verschillend kan zijn, moet de arts beschikken over diverse visualisatie opties om de voorspellingen te laten zien en te bespreken.

Het **communiceren van voorspellingen van AI-modellen** en hun implicaties voor de behandeling zal in de praktijk weerbarstig zijn. Zeker bij beslissingen in een kritische situatie is het gesprek met patiënten of naasten niet altijd mogelijk. Naasten kunnen zich overdonderd voelen en/of emotioneel niet in staat zijn om mee te beslissen. Tevens dien je als zorgverlener na te gaan of de patiënt en/of naasten de uitkomsten van voorspellende AI-modellen überhaupt willen weten. Hier ligt het recht op niet weten ten grondslag. Echter, ook in dergelijke **moeilijke situaties**, zou je als zorgverlener moeten proberen om, op zijn minst, uit te kunnen leggen aan de patiënt, waarom het voor de beslissing belangrijk is om zijn/haar voorkeuren te weten. Het is tevens van belang om de patiënten en/of naasten vanaf het begin geregeld te informeren over de voortgang en behandelopties en niet alleen vanaf het moment dat het niet goed gaat. Daarmee geeft je de patiënten of naasten de kans om te **groeien in hun rol als gesprekspartner**. Dit aspect van de arts-patiënt relatie heeft niet veel te maken met voorspellende AI-modellen, maar is cruciaal om op te kunnen terugvallen als we voorspellende AI-modellen willen gaan gebruiken ter ondersteuning van medische beslissingen en in het gesprek met onze patiënten.

In het artikel van Geurts et al. wordt een parallel getrokken tussen **critical appraisal** van wetenschappelijke literatuur door zorgverleners en hoe zorgverleners om zouden moeten gaan met uitkomsten van voorspellende modellen in het gesprek met de patiënt. De auteurs stellen voor dat, net zoals zorgverleners resultaten uit wetenschappelijke studies beoordelen op hun meerwaarde voor eigen patiënten, zorgverleners de uitkomsten van voorspellende modellen samen met patiënten kritisch zouden moeten beoordelen. Dit om potentiële biases te identificeren die van invloed zouden kunnen zijn op de medische beslissing maar ook om het gesprek aan te gaan over de onzekerheden rondom de uitkomst en persoonlijke voorkeuren van de patiënt. Hier komt weer EBM om de hoek kijken waar drie componenten noodzakelijk zijn: **evidence, expertise en voorkeuren van patiënt**. Evidence is voor beide partijen inzichtelijk (uitkomsten voorspellend model), expertise is wat de arts inbrengt, patiënt voorkeuren is wat patiënten en/of naasten inbrengen. Op die manier kun je geïndividualiseerde beslissingen nemen. Artsen en patiënten gaven aan dat, zeker in het geval van besluitvorming over wel of niet stoppen met levensbehoudende IC-behandeling, het wennen aan de prognose zowel voor artsen als naasten een proces is dat tijd kost. Daar zullen voorspellende AI-modellen niet zo veel aan veranderen. De te nemen tijd moet echter niet ten koste gaan van het welbevinden van de patiënt, omdat zinloos doorbehandelen ook ongewenst en schadelijk is.

## **6.2. Ziekenhuizen moeten transparant zijn over het gebruik van voorspellende AI-modellen ter ondersteuning van medische beslissingen.**

Over de mate van transparantie naar buiten toe die ziekenhuizen zouden moeten toepassen bij in gebruik nemen van voorspellende AI-modellen, hebben we een verschil van mening geconstateerd tussen geïnterviewde artsen en patiënten. De artsen vonden over het algemeen dat het gebruik van voorspellende AI-modellen niet per se iets is om als ziekenhuis naar buiten toe te communiceren. Net als dat je niet naar buiten toe communiceert dat je een nieuwe MRI-scanner in gebruik neemt, wat ook een vorm van technologische innovatie is. Patiënten gaven aan wel te willen weten welke voorspellende AI-modellen hun ziekenhuis gebruikt. Deze behoefte kwam vooral voort uit de behoefte om als gelijkwaardige gesprekspartner gezien te worden, dus ook kennis te hebben over welke bronnen van informatie zorgverleners gebruiken. Mogelijk ligt het ook aan de nieuwigheid van voorspellende AI-modellen dat het om meer transparantie naar buiten toe vraagt. Ondanks dit meningsverschil, hebben we ervoor gekozen om patiënten te volgen en deze eis in dit rapport op te nemen. Dit omdat nu eenmaal de patiënten van nu en van de toekomst diegenen (zullen) zijn waarop de uitkomsten van voorspellende AI-modellen van invloed (zullen) zijn.

## **SLOTWOORD**

De bevindingen uit dit onderzoek laten zien dat we, om te blijven voldoen aan de eed van Hippocrates “first do no harm”, even waakzaam en veeleisend moeten zijn voor AI-technologie als voor alle andere medische hulpmiddelen die van invloed zijn op de behandeling van onze patiënten.

Omdat de bevindingen gepresenteerd in dit rapport gebaseerd zijn op een beperkt aantal interviews, zijn de geformuleerde randvoorwaarden mogelijk minder goed generaliseerbaar naar behoeften en wensen van artsen en patiënten die niet zijn geïnterviewd. Echter, door voorafgaand aan de interviews een uitgebreide literatuurstudie te doen en door de randvoorwaarden zo veel mogelijk te verduidelijken met bevindingen uit de literatuur, zijn we van mening dat de 22 randvoorwaarden breed gedragen behoeften en wensen representeren. Het sterke punt van dit onderzoek is tevens de deelname van diverse stakeholders vanuit diverse organisaties en met verschillende achtergronden.

Reacties op dit rapport kunnen per e-mail verstuurd worden naar Dr. Joanna Klopowska op: [j.e.klopowska@amsterdamumc.nl](mailto:j.e.klopowska@amsterdamumc.nl)

## REFERENTIES

- Ahuja et al. *The impact of artificial intelligence in medicine on the future role of the physician*. PeerJ 2019; Vol.7: e7702
- Barocas S et al. *Big Data's Disparate Impact*. 104 CALIF. L. REV. 671 (2016)
- Blom Celine - Artikelen reeks van de Nederlandse Stichting voor Psychotechniek over bias mechanismen bij het ontwikkelen van AI algoritmen: <https://www.innovatiefinwerk.nl/toekomst-van-werk-diversiteit/2019/04/waarom-ai-niet-neutraal-bias-1><https://www.stichting-nice.nl/doc/jaarboek-2017-web.pdf>
- NVZ & NFU. *Convenant "Veilige Toepassing van Medische Technologie in de medisch specialistische zorg"*. Barnyard Creative Powerhouse, Utrecht 2016
- Federatie Medische Specialisten. *Themanummer De Medisch Specialist 2025*. Maart 2017
- Geurts et al. *End-of-life decisions in patients with severe acute brain injury*. Lancet Neurol 2014; Vol.13:515–24
- Gianfrancesco et al. *Potential Biases in Machine Learning Algorithms Using Electronic Health Record Data*. JAMA Intern Med. 2018; Vol. 178(11):1544-1547
- Hsu J. Medical Advice From a Bot: The Unproven Promise of Babylon Health: <https://undark.org/2019/12/09/babylon-health-artificial-intelligence-medical-advice/>
- Jobin et al. *The global landscape of AI ethics guidelines*. Nature Machine Intelligence 2019 Vol.1:389–399
- Kepecky et al. *Risk Calculators and Decision Aids are not enough for SDM*. JAMA Surg 2019; Vol.154(1):3-4.
- Kish-Gephart, J. J., Harrison, D. A. & Treviño, L. K. *Bad apples, bad cases, and bad barrels: Meta-analytic evidence about sources of unethical decisions at work*. J Appl Psychol 2010; Vol.95:1–31
- Knaus et al. *The SUPPORT prognostic model. Objective estimates of survival for seriously ill hospitalized adults. Study to understand prognoses and preferences for outcomes and risks of treatments*. Ann Intern Med. 1995; Vol.122(3):191-203.
- Mittelstadt B. *Principles alone cannot guarantee ethical AI*. Nature Machine Intelligence November 2019
- Mols B. *Internationaal AI-beleid: Domme data, slimme computers en wijze mensen*. Working Paper nummer 34. Wetenschappelijke Raad voor het Regeringsbeleid, Den Haag 2019, E-ISBN 978 94 901 86 76 0
- Neuteboom O. en Sijmons J.G. *Rapport Big Data in de Zorg: geheimhouding en privacy*. Molengraaf Instituut voor Privaatrecht. Universiteit van Utrecht, Juli 2012
- Offringa M et al. *Inleiding evidence-based medicine: klinisch handelen gebaseerd op bewijsmateriaal*. Bohn Stafleu van Loghum, 2008
- Plas van der et al. *Beyond speculative Robot Ethics*. Account Res. 2010; Vol.17(6): 299-315
- Proud et al. *Level of autonomy to design into human spaceflight vehicle*. NASA Johnson Space Center 2101 NASA Road 1, Houston, Texas, 77058, USA
- Reader et al. *Impossible decision? An investigation of risk trade-offs in the intensive care unit*. Ergonomics 2018; Vol 61(1):122-133.
- Schmidt C. *M.D. Anderson Breaks With IBM Watson, Raising Questions About Artificial Intelligence in Oncology*. J Natl Cancer Inst 2017, Vol. 109(5):4-5
- Shah et al. *Big Data and Predictive Analytics - Recalibrating Expectations*. JAMA 2018; Vol.320(1):27-28
- Shortliffe et al. *Clinical Decision Support in the Era of Artificial Intelligence*. JAMA 2018; Vol. 320(21):2199-2200.
- Soliman et al. *The ability of intensive care unit physicians to estimate long-term prognosis in survivors of critical illness*. Journal of Critical Care 2018; Vol.43:148–155
- Vaughan A. *Google gets green light to access five years of NHS patient data*. New Scientist 17 October 2019.
- Wyatt et al. *Prognostics models: clinically useful or quickly forgotten?* BMJ 1995; Vol.311:1589
- Wessler BS, Paulus JK, Lundquist C, et al. *Tufts PACE Clinical Prediction Model Registry: update 1990 through 2015*. Diagn Progn Res 2017; Vol.1(20):1-8.