

# Eerste aanzet voor een Health Technology Assessment (HTA) raamwerk voor medische Artificial Intelligence (AI): een kwalitatieve en kwantitatieve aanpak



By

Irene de Bruin

A Bachelor's Thesis

Submitted to the Department of Science & Technology

Bachelor programma Gezondheidswetenschappen

University of Twente

In fulfilment of the requirements for a Bachelor of Science (BSc) Degree

September 2024

## Samenvatting

**Inleiding en theoretisch kader:** Momenteel heeft de zorgsector te maken met discrepanties in zorgvraag en -aanbod, zorguitgaven, toegankelijkheid en betaalbaarheid van zorg. Medische Artificial Intelligence (AI) toepassingen kunnen deze problemen verminderen, omdat dit computersystemen zijn die de menselijke intelligentie kunnen nabootsen. Echter, medische AI toepassingen worden nog weinig geïmplementeerd in de zorg, omdat besluitvormers en betrokken stakeholders niet weten hoe deze complexe nieuwe zorgtechnologieën geëvalueerd kunnen worden om hun (potentiële) waarde te schatten. Voor het evalueren van innovatieve zorgtechnologieën, zoals medische AI, gaat de voorkeur naar het gebruiken van Health Technology Assessment (HTA) raamwerken, omdat deze systematisch en multidisciplinair uitgevoerd worden. De benodigde aspecten (zoals context informatie) en criteria (zoals tijdshorizon) in een HTA voor het accuraat evalueren van medische AI toepassingen worden als te onduidelijk, niet-praktijkgericht en onbekend ervaren. Dit onderzoek wil een eerste aanzet ontwerpen voor een HTA raamwerk voor de assessment van medische AI toepassingen (HTA-mAIx).

**Methodologie:** Dit exploratieve onderzoek had zowel een kwalitatief als kwantitatief onderzoeksdesign en bestond uit drie fasen. De eerste fase bestond uit semigestructureerde interviews voor het identificeren van benodigde aspecten en criteria en onderliggende fenomenen met behulp van de thematische analyse methode. In de tweede fase werd een online survey afgenomen waarin de geïdentificeerde aspecten en criteria uit de interviews werden gecategoriseerd in relevant of niet relevant met 5-puntsschaalvragen per aspect. De resultaten van fase één en twee waren de primaire input voor de derde fase waarin de eerste aanzet voor de HTA-mAIx werd ontworpen.

**Resultaten en conclusie:** Uit de antwoorden van de geïnterviewde stakeholders en de online surveyrespondenten is gebleken dat de relevante aspecten de context, de betrokken perspectieven, (klinische) effectiviteit, kosteneffectiviteit, maatschappelijke impact en (technische) algemene informatie over een medische AI toepassing zijn. De bijbehorende criteria moeten gespecificeerd worden per medische AI toepassing en per applicatiegebied, omdat medische AI te omvangrijk en te verschillend van elkaar zijn. De eerste aanzet van de HTA-mAIx is daardoor als een iteratief en incrementeel proces ontworpen met zeven basisonderdelen waarin de relevante aspecten en criteria onderverdeeld zijn en dienen als basisstructuur die verdere verdieping nodig heeft.

**Discussie:** Op basis van de resultaten van dit onderzoek in vergelijking met de resultaten van vergelijkbare onderzoeken en tot dusver nieuw gepubliceerde HTA(-gebaseerde) raamwerken voor medische AI toepassingen, lijken deze in eerste instantie overeen te komen. Echter, blijkt het doel en interpretaties van toepasbaarheid voor besluitvorming in praktijksituaties erg te verschillen. De behoefte was het creëren van een overkoepelende, algemeen geldende basisstructuur bestaande uit relevante aspecten en criteria voor medische AI in de zorgsector, die overlopen in gespecificeerde criteria voor de (klinische) effectiviteit en de kosteneffectiviteit te creëren.

## Inhoudsopgave

Samenvatting .....	1
Inleiding .....	3
Theoretisch kader .....	4
Health Technology Assessment (HTA) en AI .....	4
Methodologie .....	6
Onderzoeksdesign .....	6
Fase 1: Semigestructureerde interviews .....	7
Vorbereiding .....	7
Inclusie- en exclusiecriteria .....	7
Interviewschema .....	8
Procedure .....	8
Thematische analyse interviews.....	8
Fase 2: Online survey.....	9
Vorbereiding.....	9
Inclusiecriteria en exclusiecriteria .....	9
Procedure .....	9
Analyse online survey.....	9
Fase 3: Eerste aanzet van HTA raamwerk voor medische AI in zijn algemeenheid (HTA-mAIx).....	10
Ontwerp.....	10
Resultaten.....	10
Fase 1: Semigestructureerde interviews .....	10
Algemene kenmerken geïnterviewden.....	10
Relevante aspecten en criteria uit de interviews .....	11
Relevante aspecten voor her-evaluatie .....	13
Fase 2: Online survey.....	14
Algemene kenmerken survey respondenten .....	14
Gecategoriseerde aspecten en criteria.....	14
Context .....	18
Betrokken perspectieven .....	18
(Klinische) effectiviteit .....	18
Kosteneffectiviteit (gezondheid economische evaluaties) .....	19
Maatschappelijke impact .....	19
(Technische) algemene informatie van een medische AI toepassing.....	19
Relevante aspecten voor her-evaluatie .....	20
Fase 3: Ontwerp concept HTA-mAIx.....	20

<b>Discussie</b> .....	22
Interpretatie van de resultaten .....	22
Beperkingen van het onderzoek.....	22
Aanbevelingen voor vervolgonderzoeken .....	23
Conclusie .....	23
Referenties.....	24
Appendix.....	28
A: Overzicht tot dusver gepubliceerde HTA(-gebaseerde) raamwerken voor medische AI .....	28
B: Literatuuronderzoek opzet .....	30
C: Operationalisatieschema.....	31
D: Inclusie- en exclusiecriteria en stakeholderanalyse .....	31
E: Interviewschema .....	32
F: Informed Consent interviews .....	36
G: Online survey .....	39
H: Resultaten thematische analyses.....	39
I: Kwantitatieve en kwalitatieve surveyresultaten.....	42
J: Resultaten online survey .....	42

## Inleiding

Medische Artificial Intelligence (AI) toepassingen voor in de gezondheidszorgsector zijn momenteel veel besproken onderwerpen, omdat medische AI toepassingen de potentie hebben om wereldwijd de discrepantie tussen de zorgvraag en het zorgaanbod door onvoldoende personeelscapaciteit, schaarste aan middelen en materialen en de vergrijzing te verminderen (1-7). AI is namelijk een computersysteem dat de menselijke intelligentie nabootst door combinaties van verschillende algoritmes, zoals Artificial Neural Networks (ANN) en Machine Learning (ML). Hierdoor kan het de eigenschappen van het menselijk brein imiteren op het gebied van informatieverwerking, zoals redeneren, het zelflerend vermogen en het aanpassingsvermogen (2, 3, 6, 7). De toenemende vergrijzing bij zowel de zorgverleners als de bevolking maken dat de jaarlijkse uitstroom van zorgverleners, door bijvoorbeeld pensioen, meer is dan de instroom aan nieuwe zorgverleners. Volgens het rapport van de World Health Organisation (WHO) over de "Global strategy on human resources for health: Workforce 2030" moet wereldwijd het aantal zorgverleners tussen 2013 (43,5 miljoen) en 2030 met minimaal 55% (23,8 miljoen) stijgen om de toekomstige zorgvragen en zorglasten aan te kunnen ("Universal Healthcare Coverage, UHC") (8). Echter, de onderzoeken van de WHO en Boniol et al (8, 9) uit 2022, voorspellen dat in 2030 het tekort aan zorgverleners niet kleiner dan 10 miljoen zal worden. Daarnaast heeft de moderne geneeskunde invloed op de stijgende vergrijzing, omdat mensen steeds langer met ziektes kunnen leven, waardoor het aantal patiënten met chronische ziektes blijft stijgen (5). Het gevolg is dat het stijgende aantal patiënten met chronische ziektes verantwoordelijk is voor meer dan 75% van de zorglasten in Europa en dit jaarlijks 700 miljard euro bedragen (5). De consequenties hiervan zijn de stijgende zorgkosten voor

de betalende partijen (zoals bevolking, zorgverzekeraars, ziekenhuizen), waardoor beperkte zorgtoegankelijkheid ontstaat (2, 4).

Onderzoek van Roppelt et al (4) heeft voorspeld dat het inzetten van medische AI toepassingen wereldwijd de discrepantie tussen de zorgvraag en het zorgaanbod met tenminste 20% kunnen laten afnemen voor 2026. Onderzoek van MedTech Europe (10, 11) heeft voorspeld aan de hand van praktijksituaties uit landen waar medische AI toepassingen al wel geïntegreerd zijn, dat medische AI toepassingen in applicatiegebieden, zoals diagnostiek en behandeling, jaarlijks 380.000 tot 403.000 patiëntlevens kunnen redden. Daarnaast blijkt uit dit onderzoek dat de jaarlijkse zorguitgaven in Europa met 170,9 miljard tot 212,4 miljard euro verlaagd zouden kunnen worden door ondersteuning van medische AI toepassingen (10, 11). Tevens voorspeld het onderzoek van MedTech Europe dat het jaarlijks aantal overuren als gevolg van de stijgende administratieve lasten voor zorgverleners in Europa met 1,8 miljoen uren verminderd zouden kunnen worden door ondersteuning van medische AI (10, 11).

Ondanks de enorm hoge (potentiële) waarde van medische AI toepassingen voor de zorgsector, worden deze zorgtechnologieën nog zeer weinig geïmplementeerd door betrokken stakeholders en besluitvormers in de zorgsector (2, 4, 7, 12). Volgens meerdere onderzoeken (2, 7, 13), waaronder onderzoek van Roppelt et al (4), blijken hiervoor verschillende oorzaken te zijn, waarvan het niet aansluiten van de huidige aspecten en criteria in evaluatiemodellen en/of raamwerken belangrijk is. Voor het evalueren van innovatieve, complexe zorgtechnologieën gaat de voorkeur uit naar Health Technology Assessment (HTA) raamwerken, omdat HTA's systematisch en multidisciplinair meerdere assessments (eenmalig) uitvoeren om de waarde te schatten van deze innovatieve, complexe zorgtechnologieën, zoals medische AI (14-16). Hierdoor is het doel van dit onderzoek om systematisch, objectief en met multidisciplinaire input de relevante aspecten en criteria in kaart te brengen, om een eerste aanzet te ontwerpen voor een HTA raamwerk voor de assessment van medische AI toepassingen. In dit onderzoek zijn aspecten de benodigde elementen voor in een HTA raamwerk voor medische AI toepassingen en criteria zijn de benodigde deelonderwerpen die geëvalueerd moeten worden om de (potentiële) waarde van medische AI toepassingen te schatten. Daarnaast blijkt uit meerdere studies, waaronder die van Grutters et al (17) en ZonMw (13), dat het erg belangrijk is om de juiste stakeholders te betrekken en de juiste methoden in een HTA raamwerk te gebruiken. Wanneer dit niet gebeurt, worden innovatieve, complexe zorgtechnologieën zoals medische AI toepassingen, niet naar behoren gebruikt en wordt de (potentiële) waarde niet duidelijk voor zowel de besluitvormers, de wetenschap als de maatschappij (7, 12, 13, 17). De centrale onderzoeksvraag die hieruit volgt is: *Welke aspecten en criteria zijn relevant voor het ontwerpen van een eerste aanzet tot een HTA raamwerk voor de assessment van medische AI toepassingen?*

Om deze centrale onderzoeksvraag te beantwoorden zijn deelvragen opgesteld. Deze zijn:

1. *Wat zijn de redenen/problemen voor het niet aansluiten van de huidige, gehanteerde HTA(-gebaseerde) raamwerken bij medische AI toepassingen?*
2. *Welke stakeholders moeten betrokken worden bij het ontwikkelen van een algemene gestandaardiseerde HTA voor medische AI toepassingen?*
3. *Welke aspecten en criteria worden als relevant gezien door de (key-)stakeholders voor in een HTA raamwerk om medische AI toepassingen te kunnen evalueren?*
4. *In hoeverre is het mogelijk om één nieuwe HTA raamwerk voor de assessment van elke type medische AI toepassing breed toepasbaar te maken?*

## Theoretisch kader

### Health Technology Assessment (HTA) en AI

Volgens de WHO (14) is het doel van een HTA om op een multidisciplinaire manier de (potentiële) waarde van een innovatieve zorgtechnologie te schatten. Daarnaast heeft het als doel besluitvorming te ondersteunen en te informeren door een advies te geven hoe desbetreffende innovatieve

zorgtechnologie gebruikt kan worden in zorgsystemen wereldwijd (14). Een HTA kenmerkt zich door een transparant, (wetenschappelijk) onderbouwd en incrementeel proces waarbij (wetenschappelijk) bewijs als primaire input dient (14, 18). Het incrementele kenmerk in een HTA proces impliceert het toevoegen van nieuw gepubliceerde relevante informatie of bewijzen over de te evalueren innovatie zorgtechnologie (4, 6, 7, 18, 19).

Een HTA raamwerk bestaat uit meerdere aspecten, waaronder klinische effectiviteit, kosteneffectiviteit en ethische aspecten, waarbij meerdere disciplines betrokken zijn, zoals HTA instanties (Zorginstituut Nederland, ZIN), zorgverleners en bestuur van ziekenhuizen (14, 15, 20, 21). De klinische effectiviteitsassessment schat de mate van functioneren (bijvoorbeeld of het werkt), de (potentiële) voordelen van de zorgtechnologie en de alternatieven (20, 22). De kosteneffectiviteitsassessment bestaat uit gezondheid economische evaluaties, zoals kosten-utiliteitsanalyse, die de kosten en consequenties van een innovatieve zorgtechnologie en alternatieven schatten (2, 21, 23, 24). Volgens de WHO (14) is het doel van een HTA om op een multidisciplinaire manier de (potentiële) waarde van een innovatieve zorgtechnologie te schatten. Daarnaast heeft het als doel besluitvorming te ondersteunen en te informeren door een advies te geven hoe desbetreffende innovatieve zorgtechnologie gebruikt kan worden in zorgsystemen wereldwijd (14). Het initiëren van een HTA proces met de benodigde assessment(s) wordt gedaan met een onderzoeksvraag waarin de context duidelijk naar voren komt met behulp van systematische methoden, zoals de PICOTP (Patiënt/Populatie, Interventie/Innovatie, vergelijking (Comparison), uitkomst (Outcome), Tijdshorizon en Perspectief) (25). Vervolgens kunnen de vereiste aspecten en criteria geïdentificeerd worden onder andere uit richtlijnen en wet- en regelgeving die van toepassing zijn (16, 25). Bijvoorbeeld voor predictieve en voorspellende medische AI toepassingen kunnen de 'Leidraad voor kwalitatieve diagnostische en prognostische toepassingen van AI in de zorg' (26) en de 'AI act' (27) gebruikt worden.

Meerdere onderzoeken hebben aangetoond dat de huidige gebruikte aspecten en criteria in standaard HTA raamwerken, zoals early HTA, niet of onvoldoende geschikt zijn om de (potentiële) waarde van medische AI toepassingen te schatten (4, 7, 12, 27-29). De Hond et al (12) benadrukken het belang van implementatie aspecten, zoals benodigde kennis en gebruikersvaardigheden voor eindgebruikers, om de kwaliteit van medische AI te schatten. Dit wordt niet opgenomen in een standaard HTA raamwerk/proces (12, 14, 18). Volgens Elvidge et al (21) is het van belang dat de aanvrager(s) van de HTA (zoals een zorgverzekeraar), voorafgaand aan het HTA proces, informatie over de (technische) werkingsmechanismen en het doel van de beoogde medische AI toepassing aanleveren als onderdeel van een duidelijke PICOTP (16, 21). Tevens benadrukken Elvidge et al (21) dat het beschrijven en het aanleveren van transparante bewijzen voor de juiste tijdshorizon, inclusief het leervermogen (leer curve) van de medische AI toepassing, relevant zijn voor de effectiviteitsassessments (21). Voets et al (2) constateren dat de klinische toepasbaarheid van medische AI in de gezondheidszorg mist in een HTA om de klinische effectiviteit en de kosteneffectiviteit te kunnen schatten van medische AI toepassingen. Tevens blijken uit de onderzoeksresultaten van Voets et al (2) en Vithlani et al (29) dat de gebruikte bewijzen en de (trainings)datasets voor het evalueren van medische AI toepassingen onvoldoende transparant, reproduceerbaar en objectief zijn. In de AI Act van de EU (27) is vastgesteld dat de AI ontwikkelaar alle informatie moet aanleveren op het gebied van AI en databeleid (zoals data beveiliging en dataopslag) om een uitgebreide risico(management) assessment te maken. In huidige HTA raamwerken worden de sociale, juridische, technische en ethische aspecten en criteria in vergelijking met deze AI Act niet specifiek genoeg gedefinieerd en meegenomen, die bij medische AI toepassingen juist erg belangrijk zijn (27). Volgens onder andere Elvidge et al (21) blijken deze sociale, ethische en technische aspecten een belangrijke oorzaak te zijn voor het 'black box fenomeen' dat

zoveel mogelijk voorkomen moet worden om besluitvormers en eindgebruikers, zoals zorgverleners en patiënten, te stimuleren medische AI toepassingen te gebruiken (20, 29). Het ‘black box fenomeen’ is namelijk het niet transparant weergeven van de gebruikte input parameters die de (medische) AI algoritme(s) gebruikt om tot een beslissing te komen met als gevolg dat de eindgebruiker(s), zoals artsen, deze beslissingen niet begrijpen en niet vertrouwen(30). Kolbinger et al (7) constateren het belang van het specificeren en concretiseren van de (gezondheids)uitkomsten per applicatiegebied en per type medische AI algoritme of model. Dit voorkomt het ‘black box fenomeen’ en zorgt voor het includeren van de juiste studies om de benodigde kwalitatieve en kwantitatieve data te verzamelen voor het schatten van de klinische effectiviteit (7).

Diverse onderzoekers en overheidsinstanties hebben getracht een HTA(-gebaseerde) raamwerk te ontwerpen om de waarde van medische AI toepassingen op een juiste manier te kunnen schatten (16, 19, 20, 31-40). Echter, de omvang van deze opgave blijkt ingewikkelder te zijn dan gedacht, omdat de tot dusver gepubliceerde HTA(-gebaseerde) raamwerken voor medische AI (nog) niet alle relevante aspecten en criteria omvatten (appendix A) (25, 36, 37, 39, 40). Bijvoorbeeld, Model for ASsessing the value of Artificial Intelligence (MAS-AI) van Fasterholdt et al (34) is tot dusver alleen nog toepasbaar bij AI in medische beeldvorming in zorgsystemen vergelijkbaar met Denemarken. Jiu et al (19), Belsile-Pipon et al (16) en andere onderzoekers (13, 20, 37) hebben aangetoond dat besluitvormers in de zorg behoefte hebben aan één type (HTA) raamwerk dat wereldwijd toepasbaar is bij alle soorten medische AI toepassingen en allesomvattend is. Dit impliceert dat de MAS-AI momenteel nog niet aansluit bij deze behoeften (33, 34). Daarnaast lijkt het multidisciplinair aspect (nog) niet voldoende aanwezig te zijn bij de tot dusver gepubliceerde HTA(-gebaseerde) raamwerken voor medische AI toepassingen (35, 37-40). Bijvoorbeeld het HTA Core Model van de Europese Unie netwerk voor HTA (EUnetHTA) (35) en de Next Generation HTA (HTx) (38) zijn (nog) vanuit een overheids- of technisch perspectief ontworpen (39, 40). Verder zijn bepaalde gepubliceerde HTA raamwerken voor medische AI niet vanuit de praktijk ontworpen en/of geïntroduceerd, wat impliceert dat deze HTA raamwerken niet aansluiten bij de behoeften (zoals makkelijk in gebruik) en in de praktijk niet succesvol lijken (25). Bijvoorbeeld heeft Boverhof et al (25), een standaard HTA raamwerk uitgebreid met geïdentificeerde aspecten en criteria uit systematische reviews. Hierdoor zijn de onderdelen uit de AI Act van de EU (27) lastig te integreren zoals hier voor beschreven is.

## Methodologie

### Onderzoeksdesign

Een exploratief onderzoek met zowel een kwalitatief als kwantitatief onderzoeksdesign was opgezet en uitgevoerd om een eerste aanzet te ontwerpen voor een HTA raamwerk om de (potentiële) waarde van medische AI toepassingen op de juiste manier te schatten en evalueren (19, 41). Vanuit de principes van de Grounded Theory (GT) methode (42-44), aspecten van Human-Centred Design (HCD) (45-47) en de empirische cyclus (48), was de keuze gemaakt om interviews af te nemen en werd een online survey uitgezet. Dit stond toe om holistisch en vanuit praktijksituaties de centrale onderzoeksvraag te beantwoorden en kon een eerste aanzet voor een HTA raamwerk voor de assessment van medische AI toepassingen (hierna te noemen HTA-mAix) ontworpen worden (19, 47). De deelvragen en centrale onderzoeksvraag werden beantwoordt door drie opeenvolgende fasen. Deze drie fasen waren semigestructureerde interviews, online survey en het ontwerpen van een eerste aanzet voor de HTA-mAix. De eerste deelvraag over problemen met HTA's voor AI, de tweede deelvraag over stakeholder betrokkenheid en de derde deelvraag over relevante aspecten en criteria werden beantwoordt door de eerste twee fasen. De resultaten uit fase drie gaven antwoorden op de vierde deelvraag over de mogelijkheden voor een AI HTA raamwerk en uiteindelijk de centrale onderzoeksvraag.

## Fase 1: Semigestructureerde interviews

### Vorbereiding

Relevante aspecten en criteria voor de assessment van medische AI toepassingen om de (potentiële) waarde van medische AI toepassingen te schatten werden op een systematische manier geïdentificeerd door middel van een (kort) literatuuronderzoek (appendix B). Deze geïdentificeerde aspecten en criteria werden verwerkt in een operationalisatieschema om een theoretische basis te leggen (appendix C) (4, 49, 50). Dit diende als primaire input voor de inclusie- en exclusiecriteria, de stakeholderanalyse (51) (appendix D) en het interviewschema, omdat het operationalisatieschema impliceerde welke kennis, vaardigheden en andere expertise nodig zijn voor het beoordelen van medische AI toepassingen (4, 14, 41, 46, 49, 50, 52).

### Inclusie- en exclusiecriteria

De inclusiecriteria en exclusiecriteria voor de interviews (appendix D) werden opgedeeld in twee categorieën, namelijk 1) perspectief en 2) kennis en expertise. De categorie perspectief gaf aan welke (key-)stakeholdergroepen (perspectieven) de stakeholders moesten vertegenwoordigen om wel of niet geïnccludeerd te worden. Een stakeholder kon representatief zijn voor één of meerdere van de volgende perspectieven:

- Het medisch en klinisch perspectief, zoals medisch specialisten en verpleegkundigen (52-55).
- Het economisch en toezichhoudend perspectief, zoals overheidsinstanties als Zorginstituut Nederland (ZIN) (37, 53, 54).
- Het patiënt perspectief, zoals patiëntverenigingen en patiëntfederaties (37, 52, 53, 55).
- Het academisch perspectief, zoals onderzoekers op gebied van HTA raamwerken voor complexe innovatieve zorgtechnologieën als medische AI toepassingen (52).
- Het technologisch perspectief, zoals de ontwikkelaars van de te evalueren medische AI toepassing (37, 52, 53).
- Het organisatorisch perspectief, zoals Raad van Bestuur van een ziekenhuis (37, 52, 53, 56).

Perspectieven die werden geëxcludeerd voor de interviews waren:

- Het farmaceutisch en farmaco-economisch perspectief, zoals farmaceuten en farmaco-economen (52, 53).
- Niet-zorg gerelateerde perspectieven (52-55).

De categorie kennis en expertise was gericht op de inhoudelijke kenmerken, kennis, ervaring en/of expertise waarover een (key-)stakeholder(groep) tenminste twee moest beschikken op gebied van HTA raamwerken en/of medische AI toepassingen om geïnccludeerd te worden. De inclusiecriteria waren:

- Kennis en expertise in het ontwikkelen van AI toepassingen voor in de zorgsector (57).
- Ervaring met het evalueren, gebruiken en/of onderzoek doen naar medische AI toepassingen in het algemeen en/of binnen een specifiek applicatiegebied.
- Het kunnen uitvoeren van HTA raamwerken bij complexe innovatieve zorgtechnologieën, met name AI-gerelateerde zorgtechnologieën en/of medische AI software (52, 57).
- Kennis en expertise in het implementatieproces en/of het implementeren van complexe innovatieve zorgtechnologieën, met name AI-gerelateerde zorgtechnologieën en/of medische AI software (58).
- Kennis en ervaring met het evalueren van medische AI toepassingen met behulp van HTA raamwerk(en) en/of gezondheid economische evaluaties en/of andere evaluatiemodellen (52, 57).

- Het kunnen toepassen en/of ontwikkelen van evaluatiecriteria voor AI toepassingen in de zorgsector bij één of meerdere zorgapplicatiegebieden, zoals diagnostiek en monitoring (21, 57).
- Onafhankelijk, objectief en niet-bevooroordeeld zijn bij het evalueren van medische AI toepassingen in het algemeen en/of specifiek met HTA(-gebaseerde) raamwerken en/of gezondheid economische evaluaties (58).

De inhoudelijke kenmerken, kennis, ervaring en/of expertise die geëxcludeerd werden voor de interviews waren:

- Het ontwikkelen en/of toepassen van HTA raamwerken voor de farmaceutische industrie en andere medicamenteuze behandelingen waarbij medische AI toepassingen geen rol spelen (58).
- Het toepassen van HTA raamwerken voor ziekte preventieprogramma's waarbij medische AI toepassingen geen rol spelen (58).

### Interviewschema

De interviews waren semigestructureerd, omdat hierdoor nog niet ontdekte relevante fenomenen vanuit de praktijk voor het ontwerpen van het concept HTA-mAIx, geïdentificeerd konden worden (41, 59). Dit sluit aan bij de empirische cyclus en het multidisciplinaire perspectief van het HCD raamwerk (45, 48). Het interviewschema (appendix E) bestond uit vier onderdelen: achtergrond informatie, benodigde input/output voor het uitvoeren van een HTA voor medische AI toepassingen, randvoorwaarden en afsluiting. De inhoud van het interviewschema was gebaseerd op het operationalisatieschema. De inhoudelijke besprekpunten bij benodigde input/output waren de benodigde perspectieven, de (klinische) effectiviteit, benodigde technische informatie van medische AI en kosteneffectiviteit. De besprekpunten bij de randvoorwaarden waren her-evaluatie (re-assessment), incrementeel evalueren, stakeholder betrokkenheid, wet- en regelgeving.

### Procedure

Tussen .....(60) en ...via e-mail benaderd door de onderzoeker voor een interview. Zorginstituut Nederland (ZIN) (61) en onderzoekers op gebied van evaluatiemodellen en medische AI werden via e-mail benaderd door de eerste begeleider namens de onderzoeker voor een interview. Het streven was om minimaal vier interviews af te nemen waarbij tenminste vier van de belangrijkste perspectieven vertegenwoordigd waren. Bij bereidheid tot deelname werd een Informed Consent (IC) formulier gemaïld om vooraf het interview in te vullen met betrekking tot anonimiteit, het opnemen van het interview en informatieverwerking (appendix F) naar desbetreffende stakeholder (62).

### Thematische analyse interviews

De opnames van de interviews werden getranscribeerd met behulp van Microsoft Teams of Microsoft Word met de functie "transcribe". Vervolgens werden de opnames met bijbehorende transcripten nageluisterd op taalfouten in het transcript om deze te verbeteren. Daarna is elk getranscribeerd interview geanalyseerd volgens de thematische analysemethode, omdat deze methode aansluit bij de principes van GT door nieuwe patronen en thema's te ontdekken (41). Hierbij is zowel inductief (data-gedreven) als deductief (theorie-gebaseerd) gecodeerd en gethematiseerd (41). Het deductief coderen en thematiseren werd gedaan aan de hand van de dimensies, variabelen en/of indicatoren uit het operationalisatieschema en dienden daardoor als 'metacodes' (49). De thematische analyses werden per geïnterviewde stakeholder apart weergegeven in een tabel met in de eerste kolom de letterlijke tekst uit het transcript, in de tweede kolom een samenvatting hiervan, in de derde kolom de bijbehorende code en in de vierde kolom de inductieve en deductieve thema's (63). Hieruit werden de relevante aspecten en criteria uit de thema's geïdentificeerd en weergegeven in een tabel.

## Fase 2: Online survey

### Vorbereiding

De online survey (appendix G) werd gemaakt in Microsoft Forms op basis van de geïdentificeerde aspecten, criteria en fenomenen uit de thematisch geanalyseerde interviews en het operationalisatieschema. Deze online survey werd gecontroleerd en getest door de eerste begeleider en twee medestudenten.

De online survey bestond uit een introductie waar het doel van dit onderzoek en deze survey werden toegelicht, een invul instructie, een IC formulier en meerkeuze vragen met toelichtingsvragen na elk onderwerp bij elk inhoudelijk onderdeel. Deze inhoudelijke onderdelen waren:

- Achtergrond vragen;
- De aspecten: context en betrokken perspectieven;
- Het aspect: (klinische) effectiviteit;
- Het aspect: kosteneffectiviteit (gezondheid economische evaluaties);
- Andere relevante aspecten en criteria;
- Het ontwerp van een HTA raamwerk voor medische AI toepassingen;
- Afsluitende vragen.

Bij elk onderwerp (aspect) werden de bijbehorende criteria uit de thematische analyses en het operationalisatieschema vermeldt. De respondenten konden met een 5-puntsschaal aangeven hoe belangrijk elk criterium was voor het evalueren en schatten van de (potentiële) waarde van medische AI toepassingen. Een "1" betekende helemaal niet belangrijk en een "5" zeer belangrijk.

### Inclusiecriteria en exclusiecriteria

De inclusie- en exclusiecriteria voor het identificeren van de benodigde stakeholders en stakeholdergroepen voor de online survey waren hetzelfde als bij de interviews. De respondenten kregen de mogelijkheid om aan te geven representatief te zijn voor meerdere stakeholdergroepen.

### Procedure

De uitnodigingen voor het invullen van de online survey werden tussen ....door de onderzoeker via e-mail naar de geïnterviewde stakeholders uit fase 1 en naar de stakeholdergroepen ....verstuurd. Op .... had de eerste begeleider namens de onderzoeker ....de online survey uitgezet en uitnodigingen gemaïld naar onderzoekers op het gebied van evaluatiemodellen en/of HTA raamwerken en/of medische AI. De onderzoeker gaf toestemming aan de benaderde stakeholder(groepen) voor het delen van de uitnodiging voor de online survey met andere disciplines die ervaring, kennis en expertise hebben met HTA raamwerken of evaluatiemodellen en/of ervaring hebben met het evalueren van medische AI. Tussen .... werden herinneringse-mails verstuurd door de onderzoeker en de eerste begeleider namens de onderzoeker naar de benaderde stakeholder(groepen) en de uitnodiging voor de online survey weer ....gedeeld. Het streven was om minimaal 30 responses op de online survey te krijgen.

### Analyse online survey

De respondenten van de online survey en hun gegeven antwoorden werden anoniem door de onderzoeker ontvangen om deze stakeholders te beschermen en om bias tijdens de analyse te voorkomen (62). Er werd een overzicht gemaakt met de algemene kenmerken van de respondenten van de online survey. Deze algemene kenmerken waren onder andere het aantal respondenten (zowel algemeen als per stakeholdergroep), land van afkomst en ervaring met HTA raamwerken en medische AI toepassingen. Voor elk aspect en criterium waren gemiddelde scores berekend aan de hand van de antwoorden van de respondenten op de 5-puntsschaalvragen, die vervolgens gecategoriseerd werden in relevante en niet relevante aspecten en criteria (64). De drempelwaarde

voor het onderscheiden van de relevante en niet relevante aspecten en criteria was 3,00. Een gemiddelde score  $\geq 3,00$  toonde aan dat de respondenten een aspect of criterium relevant vinden, een gemiddelde score  $< 3,00$  toonde aan dat de respondenten een aspect of criterium niet relevant vonden. De reden voor deze drempelwaarde is dat de score 3,00 'neutraal' betekent en daardoor geen bijdrage levert bij het categoriseren (64). De gemiddelde scores per aspect en criterium met bijbehorende spreidingsmaten en de kwalitatieve data uit de toelichtingsvragen werden gebruikt om de uitkomsten context te geven door verschillen en overeenkomsten te identificeren (65). De kwalitatieve data uit de open- en toelichtingsvragen werden geanalyseerd met de thematische analyse methode, waarbij de inductieve benadering werd gebruikt om nieuwe fenomenen, andere aspecten en criteria te identificeren (39, 64, 65).

### Fase 3: Eerste aanzet van HTA raamwerk voor medische AI in zijn algemeenheid (HTA-mAIx)

De geïdentificeerde aspecten en criteria uit de resultaten van de interviews en de online survey werden gebruikt om de eerste aanzet van het HTA-mAIx te ontwerpen.

#### Ontwerp

De aanpak voor het ontwerpen van de eerste aanzet voor de HTA-mAIx was gebaseerd op de eerste drie ontwerpstadia van HCD (45, 46). Deze zijn 'early concept' waarin eerste ideeën voor prototypen geschetst werden, 'early planning' waarin een getekend prototype werd gemaakt van de ideeën en 'design refinement' waarin het getekende prototype gedigitaliseerd wordt (46). In dit onderzoek was tijdens de analyse van de interviews en de online survey al begonnen met het 'early concept' stadium door aantekeningen te maken (42-44). De 'early planning' en 'design refinement' stadia werden uitgevoerd na het voltooien van de analyse van de online survey, omdat hierdoor de informatie over de gewenste informatie architectuur (layout), aspecten en criteria bekend werden. In het 'design refinement' stadium werd de HTA-mAIx visueel gedigitaliseerd met behulp van het grafisch ontwerpprogramma Canva (66).

## Resultaten

### Fase 1: Semigestructureerde interviews

#### Algemene kenmerken geïnterviewden

Elk interview duurde gemiddeld 51 minuten, waarbij het kortste interview 23 minuten duurde en het langste interview 72 minuten. In tabel 1 toont de algemene kenmerken van de geïnterviewde stakeholders.

Tabel 1. Algemene kenmerken populatie.  $N = 5$

Geïnterviewde stakeholders	Land en/of continent	Perspectief	Ervaring met:	
			HTA	AI
<b>Anoniem #1</b>	Nederland	Academisch (onderzoeker); Medisch en klinisch; Technologisch;	Ja	Ja
<b>Anoniem #2</b>	Nederland	Economisch en toezichhoudend;	Ja	Nee
<b>Anoniem #3</b>	Nederland	Economisch en toezichhoudend; Academisch (onderzoeker);	Ja	Nee
<b>Anoniem #4</b>	Nederland; Europa;	Academisch (onderzoeker); Economisch en toezichhoudend; Medisch en klinisch;	Ja	Ja
<b>Anoniem #5</b>	Nederland; Europa;	Medisch en klinisch; Academisch (onderzoeker);	Ja	Ja

Afrika;  
United States of America  
(USA);

Economisch en toezichhoudend;

Vijf stakeholders hadden aangegeven bereid te zijn om geïnterviewd te worden en hadden het IC formulier ingevuld. De vijf interviews zijn afgenomen in het Nederlands en vier van de vijf interviews werden afgenomen, opgenomen en getranscribeerd via Microsoft Teams. Het andere interview werd op locatie afgenomen en opgenomen met een mobiel apparaat. Deze interviewopname werd met behulp van Microsoft Word getranscribeerd. Alle stakeholders vertegenwoordigden meerdere perspectieven. Drie van de vijf stakeholders vertegenwoordigden het perspectief medisch en klinisch in combinatie met academisch, waarbij één stakeholder het technologisch perspectief vertegenwoordigde gezien haar opleidingsachtergrond en ervaring met medische AI toepassingen in de beeldvormende diagnostiek. Vier van de vijf stakeholders vertegenwoordigden het economisch en toezichhoudend perspectief of waren vanuit het academisch en/of medisch en klinisch perspectief ook betrokken bij dit gebied. Alle stakeholders hadden aangegeven ervaring te hebben met HTA raamwerken, waarvan drie stakeholders ook ervaring hebben met het evalueren van medische AI toepassingen zowel met als zonder HTA raamwerken.

#### Relevante aspecten en criteria uit de interviews

Appendix H geeft een overzicht weer van de geïdentificeerde aspecten en criteria uit de vijf thematisch geanalyseerde interviews. Zes inhoudelijke aspecten voor het accuraat evalueren van medische AI applicaties kwamen naar voren uit de resultaten om de (potentiële) waarde van medische AI toepassingen te kunnen schatten.

De vijf geïnterviewde stakeholders benoemden het aspect context als een belangrijk begin voor een HTA raamwerk om de (potentiële) waarde van medische AI toepassingen te evalueren, omdat het de omvang van de benodigde HTA zal bepalen. Bijvoorbeeld een medische AI toepassing voor medische beeldvorming heeft andere parameters, zoals sensitiviteit en specificiteit, dan medische AI toepassingen voor administratieve ondersteuning. Drie stakeholders hadden aangegeven dat een passende onderzoeksvraag volgens de PICOTP relevant is, waarbij onderscheid gemaakt moet worden tussen de primaire en secundaire uitkomstmaten per applicatiegebied en per type medische AI. Daarnaast hebben de vijf stakeholders aangegeven dat het belangrijk is om in de context vroegtijdig stil te staan bij het eindgebruikersperspectief, omdat de eindgebruikers bepalen of de te evalueren medische AI toepassing succesvol en geaccepteerd zal worden. Bijvoorbeeld wanneer een medische AI voor borstkanker diagnostiek niet voldoet aan de eisen van radiologen, zoals betrouwbare transparante resultaten en makkelijk in gebruik, dan zal een radioloog aarzelend zijn over het gebruik hiervan. De criteria die hierbij als relevant naar voren kwamen waren de betrouwbaarheid, inzicht in het mandaat (digitaal toegangsgebied), benodigde gebruikerskwalificaties, -kennis en -vaardigheden, organisatorische geschiktheid en voor- en nadelen voor direct betrokken stakeholders.

De vijf stakeholders maakten duidelijk dat een multidisciplinaire aanpak bij het evalueren van medische AI toepassingen zeer relevant is, omdat door het hele HTA proces meerdere disciplines, zoals artsen en technici, nodig zijn om bijvoorbeeld de performance te beoordelen. Bijvoorbeeld de benodigde informatie over het ontwikkelingsproces en de werkingsmechanismen van een medische AI toepassing kan alleen aangeleverd worden door de ontwikkelaar(s) hiervan. De vijf stakeholders gaven aan dat het medisch, klinisch perspectief (zoals artsen en verpleegkundigen), het patiënt, cliënt perspectief en het AI ontwikkelaars perspectief belangrijk zijn. Daarnaast benoemden drie stakeholders dat het evaluator, toetsend perspectief (zoals methodologen en klinisch epidemiologen) relevant is voor bijvoorbeeld de validatie van de bewijzen om de waarde van medische AI toepassing

te evalueren. De (inter)nationale overheidsinstellingen, wet- en regelgeving perspectief werden door vier stakeholders aangegeven als relevant voor het ethisch en juridisch beschermen van bijvoorbeeld patiëntveiligheid met de Medical Device Regulations (MDR) en zorgkwaliteit met de CE certificering. Voor onder andere de financiële ondersteuning beschouwden vier stakeholders het zorgverzekeraar perspectief als relevant.

De vijf stakeholders gaven aan doordat de scope van de effecten van medische AI zo omvangrijk zijn, er niet alleen sprake zal zijn van klinische effectiviteit om de (potentiële) waarde hiervan te kunnen evalueren. Bijvoorbeeld het gebruiken van een medische AI ondersteuning bij administratieve taken kan de administratieve last verlichten, waardoor zorgverleners minder overuren maken en meer oog voor de patiënt hebben. Dit kan resulteren in het verhogen van patiënttevredenheidsscores en werktevredenheidsscores van de zorgverleners. Daarnaast beschouwden de geïnterviewde stakeholders klinische toepasbaarheid, zoals zorg verbeterend zijn en in de praktijk toepasbaar, als een zeer relevante effectiviteitscriterium. Een stakeholder gaf hiervoor als voorbeeld het AI-mind project van de EU (67). Dit project heeft namelijk als doel om vroegtijdig in de hersenen eiwitten te identificeren die een voorstadium van dementie impliceren voordat dit tot uiting komt. Maar dit project blijkt juist de kwaliteit van leven van desbetreffende patiënten te verlagen en de zorglast voor betrokken zorgverleners te verhogen door extra nazorg voor desbetreffende patiënten, waardoor het niet klinisch toepasbaar is. Verder kwam het includeren van kwalitatief en subjectief bewijs, zoals resultaten uit interview studies en meningen van potentiële eindgebruikers, als belangrijk naar voren uit de analysesresultaten. Dit heeft wederom betrekking op de omvangrijkheid van de effecten die medische AI toepassingen kunnen hebben in een applicatiegebied, zoals beeldvormende diagnostiek en thuismonitoring van patiënten met bijvoorbeeld chronische aandoeningen. Deze bewijzen moeten wel van kwalitatief hoogstaande waarde zijn, wat impliceert dat deze bewijzen transparant en passend zijn bij de scope van de onderzoeksvraag over de te evalueren medische AI toepassing.

De benodigde criteria om de kosteneffectiviteit van medische AI toepassingen te evalueren moeten voornamelijk breed interpreteerbaar zijn en specifiek vastgesteld worden per applicatiegebied en per type medische AI toepassing, volgens vijf stakeholders. Wederom heeft dit te maken met de omvangrijkheid van de verschillende effecten die medische AI toepassingen kunnen hebben. Hierdoor is het belangrijk de vastgestelde (klinische) effectiviteitscriteria zo veel mogelijk te kwantificeren zodat dit geïncorporeerd kan worden in de gezondheid economische evaluaties in bijvoorbeeld een kosten-batenanalyse. Het hebben van een tijdshorizon passend bij de leercurve en aanpassingsvermogen van de te evalueren medische AI toepassing werd als cruciaal beschouwd door drie stakeholders. Verder gaven vijf stakeholders aan dat het belangrijk is om toekomstige uitbreidingsmogelijkheden van de AI algoritmes/modellen met andere systemen, afdelingen en vergelijkbare zorginstellingen vroegtijdig mee te nemen in de kosteneffectiviteit. Eén stakeholder gaf bijvoorbeeld aan dat een medische AI toepassing, die zorg verbeterend en kostenbesparend is, bevindt zich op de Intensive Care (IC) in het Radboud UMC. Deze medische AI toepassing kan ook toegepast worden op de IC's in andere universitaire ziekenhuizen dat op nationaal-niveau financiële voordelen kan opleveren, zoals het verlagen van de jaarlijkse zorguitgaven. Daarnaast beschouwden vier stakeholders AI-gerelateerde kostencriteria, zoals kosten voor het laten draaien van de AI algoritmes/modellen, investeren in de benodigde hardware en software producten en dataopslagkosten, als relevant om te includeren in de kosteneffectiviteit. Bijvoorbeeld om een medische AI ondersteuning voor besluitvorming bij meerdere applicatiegebieden, zoals infrastructuur en resource management, 24/7 te laten draaien komen onder andere energiekosten, en kosten voor het aanschaffen of huren van servers voor dataopslag bij kijken.

Vier stakeholders gaven aan dat het van toegevoegde waarde kan zijn om de maatschappelijke impact van een medische AI toepassing te schatten, omdat een medische AI toepassing invloed kan hebben op patiëntpopulaties en de bevolking in het algemeen. Bijvoorbeeld het schatten wat het integreren van een medische AI toepassing, waarvan de algoritmes 24 uur per dag draaien, bijdraagt aan de CO2-uitstoot van een ziekenhuis en provincie met een milieu-impactanalyse.

Medische AI toepassingen worden als complex en moeilijk te begrijpen beschouwt door het ‘black box fenomeen’, waardoor het aspect (technische) algemene informatie van de te evalueren medische AI toepassing als relevant werd geacht door vier stakeholders. Criteria op het gebied van de AI algoritmes/modellen, zoals hoe het is getraind, met welke datasets het getraind was/wordt en welke factoren en databases worden gebruikt, werden als zeer relevant aangegeven. Bijvoorbeeld het onjuist trainen van AI algoritmes/modellen met niet representatieve datasets kan ervoor zorgen dat deze medische AI toepassing met onjuiste inputdata de uitkomsten formuleert of niet kan aansluiten op datasets vanuit de praktijk. Tevens noemden vier stakeholders dat risicoanalyses met risicoscores voor veiligheid in het algemeen, dataveiligheid en (in)directe cybersecurity consequenties zeer relevant zijn bij het accuraat evalueren van medische AI. Het monitoringsproces om de effectiviteit en functionaliteit van een medische AI toepassing te controleren werd als zeer belangrijk beschouwd door vijf stakeholders, omdat AI zelflerend is. Vier stakeholders gaven hiervoor als reden dat een update, nieuwe release en/of een aanpassing door dit zelflerend vermogen van AI niet altijd een toegevoegde waarde op het functioneren kan hebben.

#### Relevante aspecten voor her-evaluatie

Tabel 2 toont de relevante aspecten met betrekking tot het ontwerpen van een HTA raamwerk voor medische AI toepassingen en het her-evalueren van een medische AI toepassing na implementatie.

Tabel 2. Geïdentificeerde aspecten voor het ontwerp van een HTA voor AI en her-evaluatie. N = 5

Aspect	
1	De (klinische) effectiviteit van de medische AI toepassing
2	De kosten effectiviteit (gezondheid economische evaluaties) van de medische AI toepassing
3	De andere waarde aspecten voor de medische AI toepassing en zorginstelling
4	De gehele HTA voor desbetreffende medische AI toepassing moet her-geëvalueerd worden, waarbij gebruik wordt gemaakt van de nieuwe, meest recente wetenschappelijke bewijzen
5	Revalidatie van het huidige bewijs dat gebruikt is voor het vaststellen van de (klinische) en kosten effectiviteit van de medische AI toepassing
6	Het monitoren van de medische AI toepassingsprestaties in bijbehorende applicatiegebied (zoals het nakomen van beloftes)
7	Vaststellen van de eindgebruiker's ervaren bruikbaarheid van de medische AI toepassing

De eerste twee aspecten uit de tabel hebben betrekking op de hiervoor benoemde aspecten (klinische) effectiviteit en kosteneffectiviteit (gezondheid economische evaluaties). Het derde aspect benadrukt de incrementele aanpak, omdat vijf stakeholders benoemden dat de evaluatiecriteria per type medische AI toepassing en per applicatiegebied in een bepaalde zorginstelling kan verschillen. De vierde en vijfde aspecten zijn gericht op het continu blijven her-evalueren van medische AI toepassingen, omdat vijf stakeholders dit nodig vonden voor deze zorgtechnologieën vanwege het zelflerend en aanpassingsvermogen. De zesde en zevende aspecten uit de tabel werden door vijf stakeholders benoemd als relevant om het functioneren van desbetreffende (geïmplementeerde) medische AI toepassing en de tevredenheid van de eindgebruikers te evalueren.

## Fase 2: Online survey

In appendix G zijn de online survey weergegeven met de codeboeken met de betekenis van elke score op de 5-puntsschaal en van de stakeholdergroepen. Deze codeboeken werden gebruikt tijdens het analyseren van de kwantitatieve data.

### Algemene kenmerken survey respondenten

De survey duurde gemiddeld 44 minuten om in te vullen door de respondenten. Zeven stakeholders hebben de online survey ingevuld. Tabel 3, 4 en 5 beschrijven de algemene kenmerken van de zeven respondenten.

Tabel 3. Algemene kenmerken respondenten(1). N = 7

Stakeholdergroep	N(%)
Patiënt(verenigingen)	0(0,00)
Medisch, klinisch	0(0,00)
Overheid, toezichhoudend	2(28,57)
AI ontwikkelaars	0(0,00)
Zorgverzekeraars	1(14,29)
Besluitvorming	2(28,57)
Academisch (onderzoekers)	5(71,43)
Andere	1(14,29)

Tabel 4. Algemene kenmerken respondenten(2). N = 7

Land	N(%)
Nederland	6(85,71)
Verenigd Koninkrijk	1(14,29)

Tabel 5. Ervaring van de respondenten. N = 7

Ervaring	N(%)	N(%)	N(%)	N(%)
	Ja	nee	Wellicht	Totaal
HTA	7(100)	0(0)	-	7(100)
HTA bij medische AI toepassingen	1(14,29)	4(57,14)	2(28,57)	7(100)
Klinische effectiviteit en Kosten effectiviteit bepalen	4(57,14)	3(42,86)	0(0)	7(100)

Twee(28,57%) van de zeven respondenten hebben aangegeven elk voor drie verschillende stakeholdergroepen representatief te zijn. Eén respondent gaf aan de stakeholdergroepen zorgverzekeraar, besluitvorming en overheid, toezichhoudend te vertegenwoordigen. De andere respondent gaf aan de stakeholdergroepen academisch, besluitvorming en ANDERE: HTA senior onderzoeker te vertegenwoordigen.

### Gecategoriseerde aspecten en criteria

Tabel 6 toont de aspecten en criteria van de 5-puntsschaalvragen uit de online survey die als relevant (groen aangegeven) en niet relevant (rood aangegeven) werden beschouwd door de respondenten voor in een HTA raamwerk om de (potentiële) waarde van medische AI toepassingen accuraat te evalueren.

Tabel 6. Relevante aspecten en criteria voor in een HTA voor medische AI. N = 7

Aspect	Criterium	Gemiddelde score van de respondenten	Spreadingsmaat (min-max)
<b>Context</b>	Beschrijving van de algemene doeleinden van een medische AI toepassing	4,57	4 – 5
	Mandaat (digitaal toeganggebied) van het AI algoritme/ AI model	4,86	4 – 5
	Beschrijving van het beoogde gebruik (gebruikersaanwijzingen)	4,71	4 – 5
	Beschrijving van de benodigde kwalificaties, kennis en vaardigheden om de medische AI toepassing te gebruiken	4,00	3 – 5
	Consistentie in de prestaties	4,57	4 – 5
	Onderzoeksvraag volgens PICOTP (Patiënt/Populatie, Interventie/Innovatie, Alternatief (Comparison), uitkomsten, Tijdshorizon en Perspectieven of setting)	3,57	1 – 5
	Multidisciplinair proces	3,29	2 – 5
	Gespecificeerde primaire en secundaire (gezondheid)uitkomsten per specifieke applicatiegebied en per type medische AI toepassing	4,71	4 – 5
	Gepaste alternatieven (comparators)	4,43	3 – 5
	Beschrijving van de beoogde eindgebruikers	4,57	3 – 5
	Beschrijving van het bewijs en data die zijn gebruikt voor de ontwikkeling van de medische AI toepassing	4,71	4 – 5
	Evalueren van de organisatorische geschiktheid (zoals werkcultuur, infrastructuur)	4,00	2 – 5
	Gebruikersgeschiktheid evalueren	3,57	2 – 5
	Geschiktheid wet- en regelgeving evalueren	3,14	2 – 5
	Evalueren van de betrouwbaarheid	4,29	4 – 5
	De potentiële leercurve door het gebruik van de medische AI toepassing	4,29	4 – 5
	Identificeren welke stakeholders voordelen zullen ervaren en welke stakeholders nadelen zullen ervaren door het gebruik van medische AI toepassing	3,43	1 – 4
<b>Betrokken perspectieven</b>	Medisch, klinisch perspectief	4,57	3 – 5
	Patiënt, client perspectief	4,29	3 – 5
	Evaluator, toetsend perspectief	4,14	3 – 5
	AI ontwikkelaars perspectief	3,57	1 – 5
	(Inter)nationale overheidsinstanties, wet- en regelgeving perspectief	3,43	1 – 5
Zorgverzekeraar perspectief	3,00	1 – 4	
<b>(Klinische) effectiviteit</b>	Klinische toepasbaarheid (zoals de bruikbaarheid van de medische AI in de praktijk)	5,00	5 – 5
	Impact op (klinische) workflow	4,14	2 – 5

	Eindgebruikers acceptatie van de medische AI toepassing bepalen	3,71	1 – 5
	Impact op (klinische) werklust en werkuren	4,57	4 – 5
	Impact op toegankelijkheid van zorg	3,71	2 – 5
	Impact op kwaliteit van leven (QoL) van de betrokken stakeholders	4,71	4 – 5
	Impact op werktevredenheid scores van zorgverleners en/of andere direct betrokken stakeholders	3,86	3 – 4
	Impact op patiënttevredenheid scores	3,71	2 – 5
	Bepalen van (klinische) effectiviteit drempelwaarden	5,00	5 – 5
	Kans op falen (kans dat de medische AI toepassing niet functioneert en mogelijk schade aanbrengt)	4,29	3 – 5
	Identificeren van andere voor- en nadelen	3,71	2 – 5
	Identificeren van alle mogelijke gezondheidsrisico's van de medische AI toepassing voor de eindgebruikers andere direct betrokken stakeholders	4,14	2 – 5
	Kwalitatief bewijs (bewijs moet kwalitatief goed zijn om geïnccludeerd te worden)	3,57	2 – 5
	Subjectief bewijs (includeren van bijvoorbeeld meningen van zorgverleners en patiënten)	3,86	2 – 5
<b>Kosten effectiviteit (gezondheid economische evaluaties)</b>	Zo veel mogelijk includeren en kwantificeren van de effectiviteitscriteria uit het aspect (klinische) effectiviteit	5,00	5 – 5
	Juiste gezondheid economische evaluaties gebruiken	4,71	4 – 5
	Juiste tijdschors (hoeveelheid tijd die benodigd is om de effectiviteit te kunnen schatten)	4,14	2 – 5
	Mogelijkheid om de medische AI toepassing uit te breiden	3,14	1 – 5
	Impact op huidige en toekomstige zorglasten	4,14	3 – 5
	Directe en indirecte (verwachte) consequenties op zorguitgaven voor desbetreffende zorginstelling (zoals ziekenhuizen)	4,57	4 – 5
	Budget-impact van het gebruik van de medische AI toepassing (voor de zorginstelling en nationale zorguitgaven)	4,43	4 – 5
	Impact op de betaalbaarheid van zorg	4,57	4 – 5
	Impact op arbeidskosten	4,43	4 – 5
	Investeringskosten voor het gebruik van de medische AI toepassing	4,71	4 – 5
	Alle kosten, inclusief verzekeringskosten, van de beoogde applicatie gebied	4,14	1 – 5
	Het vastgestelde tarief voor het bekostigen van de medische AI applicatie	3,14	1 – 5

	Bepalen van nieuwe Willingness To Pay (WTP) drempelwaarde(n) voor medische AI toepassingen (voor zorgverzekeraars, overheidsinstanties en betalende partijen)	1,57	1 – 3
	Kosten voor dataopslag	3,86	2 – 5
	Identificeren van potentiële subsidies voor de medische AI toepassing	1,86	1 – 3
	Kosten voor het laten draaien van de AI algoritme(s)/ AI model(len)	4,14	2 – 5
<b>Maatschappelijke impact</b>	Impact op kwaliteit van zorg op een nationaal niveau	3,86	2 – 5
	Impact van de medische AI toepassing op de socio-economische status van de betrokken patiëntpopulatie(s) en andere betrokken stakeholders	3,29	1 – 5
	Indien nodig, (zoals e-health AI), impact en toepassingsmogelijkheden bij niet-digitaal vaardige patiëntpopulaties wanneer de medische AI toepassing “fysieke” zorgpaden vervangt	3,86	2 – 5
	Milieu impact van de medische AI toepassing	3,43	1 – 5
	(Inter)nationaal vaststellen van “goed gebruik van medische AI toepassingen” criteria en procedures om de bruikbaarheid te kunnen evalueren	3,57	1 – 5
	Directe en indirecte ethische consequenties voor eindgebruikers en andere betrokken stakeholders identificeren	3,86	2 – 5
<b>(Technische) algemene informatie van een medische AI toepassing</b>	Uitlegbaarheid (explainability: mechanisms or operating systems of the medical AI application need to be understandable and clear for the end users)	3,29	1 – 5
	Beschrijving van de gebruikte factoren of parameters in de AI algoritme(s)/ AI model(len)	3,71	1 – 5
	Transparantie over alle informatie met betrekking tot de AI algoritme(s)/ AI model(len)	4,00	4 – 5
	AI algoritme(s) en/of AI model(len) moeten deelbaar zijn met vergelijkbare zorginstellingen	3,29	1 – 5
	Beschrijving van de representativiteit van de gebruikte trainingsdatasets voor de AI algoritme	4,14	2 – 5
	Beschrijving van de onderlinge afhankelijkheden tussen databases die de medische AI toepassing nodig heeft om te functioneren	3,86	1 – 5
	Beschrijving van de betrokkenheid van de eindgebruikers tijdens de ontwikkeling van de medische AI toepassing	3,71	1 – 5
	Benodigde implementatiestrategie per applicatiegebied en per type medische AI toepassing voor een succesvolle implementatie	3,57	1 – 5

Monitoringsproces voor de medische AI toepassing om de voortgang, effectiviteit en accuraatheid te evalueren	4,14	1 – 5
Leercurve van de medische AI toepassing	3,29	1 – 5
Evalueren/beoordelen van nieuwe updates of releases voor de medische AI toepassing	4,29	4 – 5
Beschrijving van de gebruikte trainingsmethodes voor de medische AI toepassing (zoals 'Federated learning' en 'Privacy Preserved learning')	3,14	1 – 5
Uitbreidings- en/of samenvoegingsmogelijkheden van de AI algoritmes en AI modellen met andere systemen (zoals Elektronische Patiënt Dossier)	3,57	1 – 5
Vaststellen van alle mogelijke data veiligheidsrisico's van de medische AI toepassing met behulp van risicoanalyses, inclusief risicoscores per geïdentificeerde risico	3,14	1 – 5
Vaststellen van alle mogelijke veiligheidsrisico's van de medische AI toepassing met behulp van risicoanalyses en veiligheidsassessments, inclusief risicoscores per geïdentificeerde risico	3,43	1 – 5
Directe en indirecte (potentiële) cybersecurity consequenties bij gebruik van de medische AI toepassing	3,29	1 – 5

### Context

De criteria behorende bij het aspect context hebben allemaal een gemiddelde score hoger dan 3,00. Uit de toelichtingsvraag kwam naar voren dat alle criteria relevant zijn, maar deze criteria niet allemaal van toepassing zijn voor in een HTA en dat een HTA meer gericht moet zijn op de (kosten)effectiviteit in de praktijk. Drie respondenten gaven aan een aantal criteria te missen voor de HTA context bij een medische AI toepassing. Deze criteria zijn: "[...] whether the intended AI is useful and applicable within current care.", "[...] focused on the integration process of the AI [...]", "[...]the estimation of the impact of the nature speed and effectiveness of the learning curve[...]" en "What's the room for improvement in the current health care setting?" (appendix I).

### Betrokken perspectieven

De zes benoemde perspectieven bij het aspect betrokken perspectieven hadden een gemiddelde score boven de 3,00 en zijn geïncorporeerd als relevante criteria voor dit aspect. In de bijbehorende toelichtingsvraag gaf een respondent hier over het volgende aan als redenatie: "all stakeholders should have a same idea about what the hoped benefit of the medical AI is; as soon as that's the case it doesn't matter too much anymore what stakeholders are consulted in the actual HTA phase". Tevens gaf een stakeholder aan dat het patiënt perspectief zeer relevant is, maar dat patiënten (representatieven) waarschijnlijk gelimiteerde kennis over AI hebben (appendix I).

### (Klinische) effectiviteit

De geïdentificeerde criteria voor het aspect (klinische) effectiviteit hebben allemaal een gemiddelde score hoger dan 3,00. Eén respondent beschreef in de toelichtingsvraag dat alle benoemde criteria relevant zijn, maar deze criteria niet allemaal zijn geschikt voor in een HTA. Een andere respondent

gaf aan dat het bepalen van de kans van falen en nadelen niet relevant zijn als de klinische toepasbaarheid en effectiviteit juist geëvalueerd zijn. Eén respondent gaf als toelichting voor de relevantie van kwalitatief bewijs het volgende: *“Clinical applicability and the willingness of users are the most important. Qualitative evidence usually doesn’t completely reflect that, although it can of course not be forgotten completely”*. Een respondent gaf aan de volgende criteria te missen voor (klinische) effectiviteit: *“Impact of learning curves, not just that of the AI application but that of all stakeholders involved.”* En *“Likewise if a system never reaches full optimal impact then evaluations should incorporate the (limited) efficacy”* (appendix I).

#### Kosteneffectiviteit (gezondheid economische evaluaties)

Bij de vraag over het belang van het uitvoeren van een kosteneffectiviteitsassessment voordat een medische AI toepassing geïmplementeerd wordt beantwoorden vijf(71,43%) van de zeven respondenten de vraag met ja en twee(28,57%) respondenten met nee (appendix J). Een respondent gaf hier voor als reden: *“yes before scaling up we should and no because we simply cannot assess cost effectiveness in an appropriate way if we do not measure it in practice, when the medical AI is being used in the clinical setting since it is only then that we can see what the true impact is and can be for the technology.”* (appendix I). Voor het aspect kosteneffectiviteit (gezondheid economische evaluaties) hebben twee criteria een gemiddelde score lager dan 3,00 gekregen, namelijk het bepalen van nieuwe WTP drempelwaarde(n) voor medische AI toepassingen en het identificeren van potentiële subsidies voor de medische AI toepassingen. De spreidingsmaten laten zien dat de respondenten deze twee criteria als niet relevant of neutraal beschouwden. Een respondent gaf als redenatie voor het niet relevant zijn van potentiële subsidies: *“potential funds like subsidies should only be considered when the business case has already shown positive”*. Twee respondenten gaven aan nog criteria te missen voor kosteneffectiviteit, namelijk *“[...]specification of the counterfactual[...]”, “[...]the comparators[...]”, “[...]CE perspective under a fixed budget[...]”, “[...]heterogeneity in outcomes[...]”* en *“[...]uncertainty criteria of medical AI[...]”* (appendix I). Met *“counterfactual”* worden hypothetische scenario’s bedoeld (68).

#### Maatschappelijke impact

De spreidingsmaten laten verschillen zien in welke criteria als wel of niet relevant worden beschouwd. Bijvoorbeeld het criterium *“impact en toepassingsmogelijkheden bij niet-digitaal vaardige patiëntpopulaties wanneer de medische AI toepassing “fysieke” zorgpaden vervangt”* heeft van twee(28,57%) respondenten een minimum score gekregen van 2,00 en van drie(42,86%) respondenten een maximum score van 5,00. Uit de antwoorden van de respondenten bij de bijbehorende toelichtingsvraag kwam naar voren dat de patiënt als stakeholder en ethische aspecten bij maatschappelijke impact missen. Tevens beschouwde één respondent alle benoemde criteria voor maatschappelijke impact als relevant, maar niet volledig geschikt voor in een HTA: *“again, all aspects (especially applicability to populations that are less digitally able; and the ethical aspects) are very important, but not everything should be covered in an HTA to my opinion”* (appendix I).

#### (Technische) algemene informatie van een medische AI toepassing

De spreidingsmaten per criterium laten grote verschillen zien, deze op drie criteria na een minimum score van 1,00 kregen en een maximum score van 5,00. Uit de antwoorden van de toelichtingsvraag blijkt dat twee respondenten de benoemde criteria relevant vinden, maar dat niet alle criteria geschikt zijn in een HTA. Een andere respondent beschouwde het monitoren van een medische AI toepassing als meest relevant, waarbij de andere criteria als extra’s gezien kunnen worden (appendix I).

## Relevante aspecten voor her-evaluatie

Tabel 7 laat de relevante aspecten voor her-evaluatie op basis van de gemiddelde scores zien.

Tabel 7. Relevante aspecten voor het ontwerp van een HTA raamwerk voor AI en her-evaluatie. N = 7

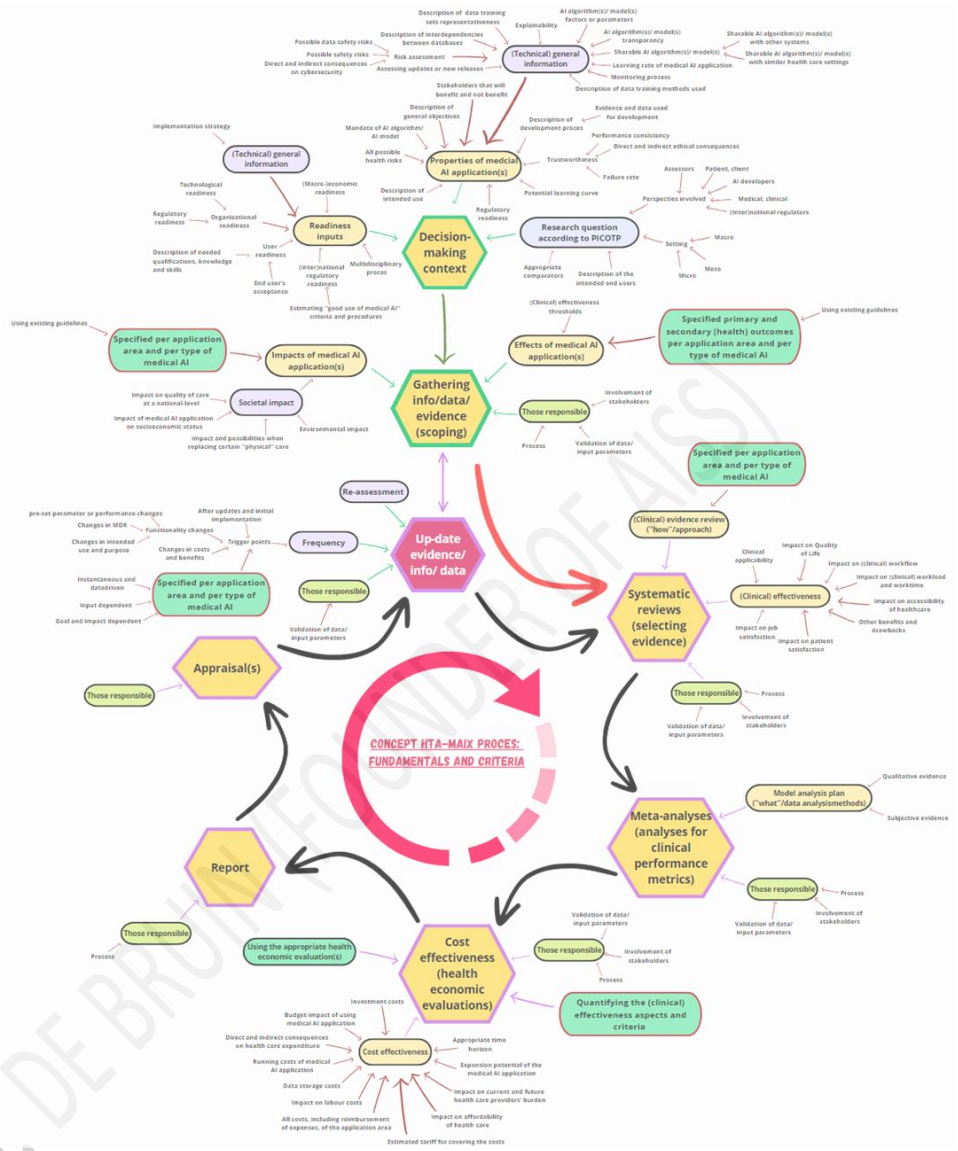
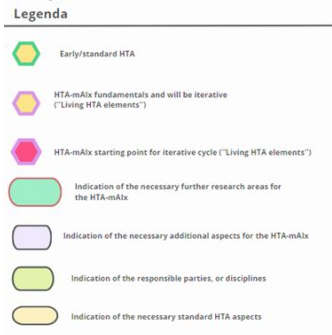
Aspect	Gemiddelde score van de respondenten	Minimum spreidingsmaat – maximum spreidingsmaat
1 De (klinische) effectiviteit van de medische AI toepassing	4,29	2 – 5
2 De kosten effectiviteit (gezondheid economische evaluaties) van de medische AI toepassing	4,57	3 – 5
3 De andere waarde aspecten voor de medische AI toepassing en zorginstelling	3,14	1 – 5
4 De gehele HTA voor desbetreffende medische AI toepassing moet re-assessed worden, waarbij gebruik wordt gemaakt van de nieuwe, meest recente wetenschappelijke bewijzen	3,57	3 – 5
5 Revalidatie van het huidige bewijs dat gebruikt is voor het vaststellen van de (klinische) en kosten effectiviteit van de medische AI toepassing	4,43	3 – 5
6 Het monitoren van de medische AI toepassingsprestaties in bijbehorende applicatiegebied (zoals het nakomen van beloftes)	4,57	4 – 5
7 Vaststellen van de eindgebruiker's ervaren bruikbaarheid van de medische AI toepassing	3,57	3 – 5

In de spreidingsmaten van de eerste en derde aspecten tonen verschillen aan in hoe relevant de respondenten deze aspecten ervaren. Uit de antwoorden van bijbehorende toelichtingsvraag kwam naar voren dat twee respondenten de relevantie van her-evaluatie afhankelijk vinden van de AI toepassing en van gebeurtenissen uit het verleden. Eén respondent gaf aan ethische implicaties voor implementatie te missen als aspect voor het evalueren van medische AI toepassingen (appendix I).

## Fase 3: Ontwerp concept HTA-mAIx

Figuur 1 toont de eerste aanzet voor de HTA-mAIx dat gebaseerd is op de resultaten van de interviews en de online survey.

Figuur 1. Eerste aanzet voor de HTA-mAIx.



Het HTA-mAix proces bestaat uit zeven basisonderdelen (*fundamentals*), waarbij de (klinische) effectiviteitsassessment is opgedeeld in systematische reviews en meta-analyses, omdat het voor medische AI toepassingen belangrijk is deze onderdelen apart te benaderen. Idealiter begint het HTA-mAix proces bovenaan bij de besluitvormingscontext (*decision-making context*) waar het doel en achtergrond duidelijk worden, zoals de onderzoeksvraag volgens PICOTP en benodigheden voor de medische AI toepassing en alternatieven (*comparators (readiness inputs)*). Vervolgens worden de bewijzen en (gezondheid)uitkomsten, effecten en impacts per applicatiegebied en per type medische AI toepassing verzameld, waaronder de maatschappelijke impact (*societal impact*). Daarna volgt via de rode pijl de systematische reviews (*systematic reviews*) waarin de benodigde (klinische) bewijzen en wetenschappelijke onderzoeken worden geselecteerd en gevalideerd. Hierbij dienen het aspect (klinische) effectiviteit (*clinical effectiveness*) met de bijbehorende criteria uit de voorgaande basisonderdelen over de context als basis uitgangspunten voor alle typen medische AI toepassingen en applicatiegebieden. Vervolgens worden de geselecteerde bewijzen en wetenschappelijke onderzoeken statistisch geanalyseerd met meta-analyses om bruikbare kwantitatieve data en input te creëren voor de kosteneffectiviteitsassessment (gezondheid economische evaluaties). Het aspect kosteneffectiviteit (*cost effectiveness*) met bijbehorende criteria dienen, net als bij de (klinische) effectiviteitsassessment, als basis uitgangspunten voor alle typen medische AI en applicatiegebieden. Daarna worden de belangrijkste uitkomsten in een rapport beschreven die als onderbouwing dienen

voor het appraisal naar de organisatie die de geëvalueerde medische AI toepassing mogelijk wil implementeren. Deze eerste aanzet voor de HTA-mAIx is ontworpen als een iteratief en cyclisch raamwerk waarmee idealiter incrementele her-evaluaties gedaan kunnen worden bij een medische AI toepassing die (mogelijk) geïmplementeerd wordt of nog verder ontwikkeld moet worden. Door het paars-roze omliggende basisonderdeel (update evidence/info/data) worden de meest recente bewijzen van desbetreffende medische AI toepassing gebruikt bij her-evaluaties.

## Discussie

### Interpretatie van de resultaten

In dit onderzoek, met zowel kwalitatieve als kwantitatieve onderzoeksmethoden, is op een systematische en holistische manier onderzocht welke aspecten en criteria relevant zijn voor het accuraat evalueren van medische AI voor in een eerste aanzet voor de HTA-mAIx. Uit de resultaten van de thematisch geanalyseerde interviews en de survey responses bleek dat het lastig is om alle relevante aspecten en criteria hetzelfde te laten zijn voor alle typen medische AI toepassingen. De resultaten tonen wel aan dat het mogelijk is om overkoepelende relevante aspecten en criteria te gebruiken voor alle typen medische AI toepassingen waarmee een basisstructuur voor de eerste aanzet van de HTA-mAIx ontworpen is. Deze aspecten zijn de context, welke perspectieven betrokken zijn, (klinische) effectiviteit, kosten effectiviteit, maatschappelijke impact en (technische) algemene informatie over de te evalueren medische AI toepassing.

De resultaten van dit onderzoek lijken in eerste instantie vergelijkbaar met resultaten van vergelijkbare onderzoeken en andere tot dusver gepubliceerde nieuw ontworpen HTA(-gebaseerde) raamwerken voor medische AI toepassingen, zoals MAS-AI (34). Het onderzoek van Boverhof et al (25), de MAS-AI (33, 34) en de Consolidated Health Economic Evaluation Reporting Standards – AI (CHEERS-AI) (21) concluderen namelijk het belang van soortgelijke relevante aspecten en criteria. Echter, het doel en interpretaties van toepasbaarheid voor besluitvorming in de praktijk met deze relevante aspecten en criteria sluiten niet aan bij de resultaten van dit onderzoek. De CHEERS-AI (21) lijkt alleen oppervlakkig onderscheid te maken in het applicatiegebied of de AI (software) en apparatuur (hardware) combinatie waarin niet specifiek wordt gekeken om wat voor type medische AI toepassing het gaat. Tijdens de interviews werd aangegeven dat het niet mogelijk is om dezelfde criteria te stellen voor alle typen medische AI toepassingen in combinatie met de verschillende applicatiegebieden. Dit heeft voornamelijk te maken met de benodigde (gezondheids)uitkomsten en de algoritme parameters passend bij de specifieke doelstelling van een medische AI toepassing (69). De MAS-AI (33, 34) is momenteel alleen nog van toepassing op medische AI toepassingen in het applicatiegebied medische beeldvormingsdiagnostiek, omdat het specifiek voor die applicatiegebied is opgesteld en wordt daarna uitgebreid naar de andere applicatiegebieden. Uit de interviews en surveyresponses blijkt dat de praktijk juist de behoefte heeft aan het tegenovergestelde, namelijk dat vanuit een overkoepelende, brede benadering het (HTA) raamwerk over gaat in gespecificeerde criteria bij desbetreffende medische AI toepassing en applicatiegebied.

### Beperkingen van het onderzoek

Dit exploratief onderzoek kent een aantal beperkingen. Allereerst, de key-stakeholdergroep patiënt is niet vertegenwoordigd. Dit was helaas ook het geval bij de AI ontwikkelaars, zoals Microsoft Nuance (6, 21, 26). Het aantal geïnterviewde key-stakeholders was echter ook gedeeltelijk representatief voor het ontwikkelaars perspectief, waardoor in fase 1 alleen het patiënt perspectief niet vertegenwoordigd is. Het gevolg is dat het doel om de HTA-mAIx vanaf het begin holistisch en multidisciplinair te ontwerpen niet gerealiseerd is en vervolgonderzoek nog meer de focus hierop moet leggen om de HTA-mAIx uiteindelijk succesvol te laten zijn (19, 28). Daarnaast heeft de online

survey erg weinig respons opgeleverd, namelijk 7 in plaats van het gestelde doel van circa 30. De oorzaken hiervan bleken vooral te liggen aan de tijdsperiode waarin de online survey was uitgezet, omdat een aantal stakeholders aangaf op vakantie te zijn. Een andere beperking van dit onderzoek is dat er geen evaluatie en validatie van de eerste aanzet HTA-mAix (figuur 1, appendix I) door de betrokken stakeholders heeft kunnen plaatsvinden, wat een cruciaal onderdeel is van zowel GT, empirische cyclus en HCD raamwerk (42, 43, 46, 48). Als laatste beperking is dat het invullen van de online survey als erg lang werd ervaren door de respondenten, omdat er heel veel aspecten en criteria in werden benoemd. Het gevolg van de benoemde beperkingen betekent dat de betrouwbaarheid, validiteit en resultaten van dit exploratief onderzoek verminderd zijn (62, 64). Dit alles leidt tot het algemeen advies voor het herhalen van dit onderzoek en vervolgonderzoeken om te beginnen op een nationaal en/of supranationaal (zoals EU) niveau stakeholders te betrekken, per basisonderdeel (*fundamental*) van de HTA-mAix te onderzoeken en dit exploratief onderzoek als een pilot te beschouwen (19).

### Aanbevelingen voor vervolgonderzoeken

Op basis van de analyseresultaten van de interviews en survey, de literatuur en de drie onderliggende methodes (GT, empirische cyclus en HCD) van dit onderzoek zijn de aanbevelingen voor vervolgonderzoeken opgesteld. De eerste aanbeveling is om het literatuuronderzoek, de interviews en/of de online survey van dit onderzoek te herhalen om meer input vanuit de praktijk te verzamelen, alle key-stakeholdergroepen te betrekken en de eerste aanzet van het HTA-mAix te verifiëren. Hierdoor zullen de resultaten representatiever zijn en een hogere mate van betrouwbaarheid en validiteit hebben (62, 70). Tevens wordt hierbij aanbevolen om dit te doen na elk vervolgonderzoek om de HTA-mAix toepasbaar en geschikt te kunnen maken voor de praktijk en uiteindelijk succesvol te implementeren.

De tweede aanbeveling is een exploratief vervolgonderzoek naar het identificeren van de (gezondheid)uitkomsten en andere parameters per applicatiegebied en per type medische AI om de HTA-mAix gedetailleerder te maken door hiervan aparte modules te creëren (42, 71).

De derde aanbeveling is het uitvoeren van een toetsend vervolgonderzoek om de, op dat moment door key-stakeholdergroepen geverifieerde, digitale concept HTA-mAix in praktijksituaties te testen middels pilots, multidisciplinaire focusgroepen en/of *cognitive walkthroughs* om het gebruiksgemak te toetsen (21, 46). Vervolgens is het belangrijk om het testen, evalueren, aanpassen en verifiëren van de HTA-mAix in de praktijk iteratief te blijven doen om de toepasselijkheid, gebruiksgemak en effectiviteit te monitoren. Bijlage 3 toont een schematisch overzicht van de benoemde aanbevelingen op chronologische volgorde vanaf de tweede benoemde aanbevelingen.

### Conclusie

Gedurende de interviews met stakeholder vanuit de praktijk en uit de antwoorden van de online survey respondenten kwam het belang voor een overkoepelende HTA raamwerk voor het accuraat evalueren van medische AI toepassingen als een zeer relevante behoefte naar voren. Uit de resultaten van de interviews en de online survey kan geconcludeerd worden dat de relevante aspecten de context, welke perspectieven betrokken zijn, (klinische) effectiviteit, kosten effectiviteit, maatschappelijke impact en (technische) algemene informatie over de te evalueren medische AI toepassing zijn. Wat betreft de relevante criteria kan geconcludeerd worden dat hiervoor aparte 'modules' of groeperingen benodigd zijn per applicatiegebied, zoals diagnostiek, en per type medische AI toepassing. Medische AI toepassingen zijn te omvangrijk en te verschillend door hun aanpassings- en zelflerend vermogen. Het is wel mogelijk om met de aspecten een overkoepelende HTA raamwerk te ontwerpen die dienen als een basisstructuur voor dit is de concept HTA-mAix.

## Referenties

1. Lekadir K. QG, Garmendia A. T., Gallin C., . Artificial Intelligence in Healthcare. Brussels: Scientific Foresight Unit (STOA); 2022. Report No.: PE 729.512.
2. Voets MM, Veltman J, Slump CH, Siesling S, Koffijberg H. Systematic Review of Health Economic Evaluations Focused on Artificial Intelligence in Healthcare: The Tortoise and the Cheetah. *Value in Health*. 2022;25(3):340-9.
3. Vassolo RS, Mac Cawley AF, Tortorella GL, Fogliatto FS, Tlapa D, Narayanamurthy G. Hospital Investment Decisions in Healthcare 4.0 Technologies: Scoping Review and Framework for Exploring Challenges, Trends, and Research Directions. *J Med Internet Res*. 2021;23(8):e27571.
4. Roppelt JS, Kanbach DK, Kraus S. Artificial intelligence in healthcare institutions: A systematic literature review on influencing factors. *Technology in Society*. 2024;76:102443.
5. Klumpp M, Hintze M, Immonen M, Ródenas-Rigla F, Pilati F, Aparicio-Martínez F, et al. Artificial Intelligence for Hospital Health Care: Application Cases and Answers to Challenges in European Hospitals. *Healthcare (Basel)*. 2021;9(8).
6. Shevtsova D, Ahmed A, Boot IWA, Sanges C, Hudecek M, Jacobs JLL, et al. Trust in and Acceptance of Artificial Intelligence Applications in Medicine: Mixed Methods Study. *JMIR Hum Factors*. 2024;11:e47031.
7. Kolbinger FR, Veldhuizen GP, Zhu J, Truhn D, Kather JN. Reporting guidelines in medical artificial intelligence: a systematic review and meta-analysis. *Communications Medicine*. 2024;4(1):71.
8. World Health Organization. Global strategy on human resources for health: Workforce 2030. Geneva, Switzerland: WHO Press; 2016. 64 p.
9. Boniol M, Kunjumen T, Nair TS, Siyam A, Campbell J, Diallo K. The global health workforce stock and distribution in 2020 and 2030: a threat to equity and 'universal' health coverage? *BMJ Global Health*. 2022;7(6):e009316.
10. MedTech Europe. Artificial Intelligence: MedTech Europe; n.d. [Available from: <https://www.medtecheurope.org/digital-health/artificial-intelligence/>]
11. Biundo E, Pease A, Segers K, de Groote M, d'Argent T, de Schaetzen E, . The socio-economic impact of AI in healthcare. European Union; 2020.
12. de Hond AAH, Leeuwenberg AM, Hooft L, Kant IMJ, Nijman SWJ, van Os HJA, et al. Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review. *npj Digital Medicine*. 2022;5(1):2.
13. ZonMw Methodologie n.d. [Available from: <https://www.zonmw.nl/nl/methodologie>].
14. World Health Organization Health Technology Assessment: World Health Organization; n.d. [Available from: [https://www.who.int/health-topics/health-technology-assessment#tab=tab\\_1](https://www.who.int/health-topics/health-technology-assessment#tab=tab_1)].
15. Institute for Clinical and Economic Review Guide to Understanding Health Technology Assessment (HTA). Institute for Clinical and Economic Review (ICER); 2018.
16. Bélisle-Pipon J-C, Couture V, Roy M-C, Ganache I, Goetghebeur M, Cohen IG. What Makes Artificial Intelligence Exceptional in Health Technology Assessment? *Frontiers in Artificial Intelligence*. 2021;4.
17. Grutters JPC, Kluytmans A, van der Wilt GJ, Tummers M. Methods for Early Assessment of the Societal Value of Health Technologies: A Scoping Review and Proposal for Classification. *Value in Health*. 2022;25(7):1227-34.
18. Ming J, He Y, Yang Y, Hu M, Zhao X, Liu J, et al. Health technology assessment of medical devices: current landscape, challenges, and a way forward. *Cost Effectiveness and Resource Allocation*. 2022;20(1):54.
19. Jiu L, Hogervorst MA, Vreman RA, Mantel-Teeuwisse AK, Goettsch WG. Understanding innovation of health technology assessment methods: the IHTAM framework. *International Journal of Technology Assessment in Health Care*. 2022;38(1):e16.

20. Thokala P, Srivastava T, Smith R, Ren S, Whittington MD, Elvidge J, et al. Living Health Technology Assessment: Issues, Challenges and Opportunities. *Pharmacoeconomics*. 2023;41(3):227-37.
21. Elvidge J, Hawksworth C, Avşar TS, Zemplenyi A, Chalkidou A, Petrou S, et al. Consolidated Health Economic Evaluation Reporting Standards for Interventions That Use Artificial Intelligence (CHEERS-AI). *Value in Health*. 2024.
22. Fattore G, Maniadakis N, Mantovani LG, Boriani G. Health technology assessment: what is it? Current status and perspectives in the field of electrophysiology. *EP Europace*. 2011;13(suppl\_2):ii49-ii53.
23. Husereau D, Drummond M, Augustovski F, de Bekker-Grob E, Briggs AH, Carswell C, et al. Consolidated Health Economic Evaluation Reporting Standards 2022 (CHEERS 2022) statement: updated reporting guidance for health economic evaluations. *MDM Policy & Practice*. 2022;7(1):23814683211061097.
24. Drummond M F. SMJ, Claxton K., Stoddart G L., Torrance G W., . *Methods for the Economic Evaluation of Health Care Programmes* Fourth ed. United Kingdom: Oxford University Press; 2015.
25. Boverhof B-J, Redekop WK, Visser JJ, Uyl-de Groot CA, Rutten-van Mülken MPMH. Broadening the HTA of medical AI: A review of the literature to inform a tailored approach. *Health Policy and Technology*. 2024;13(2):100868.
26. van Smeden M, Moons C, , Hooft L, , Kant I, , van Os H, , Chavannes N, , . Leidraad kwaliteit AI in de zorg. In: Ministerie van Volksgezondheid WeS, editor. OSFHome: OSFHome; 2023. p. 81.
27. European Commission. EU Artificial Intelligence Act. In: Commission E, editor.: European Union; 2024.
28. Belisle-Pipon J. C. CV, Roy M. C., Ganache I., Goetghebeur M., Cohen I. G., . Wat makes Artificial Intelligence exceptional in Health Technology Assessment? *Frontiers in Artificial Intelligence*. 2021;4.
29. Vithlani J, Hawksworth C, Elvidge J, Ayiku L, Dawoud D. Economic evaluations of artificial intelligence-based healthcare interventions: a systematic literature review of best practices in their conduct and reporting. *Frontiers in Pharmacology*. 2023;14:1220950.
30. Bose S. The 'black box' Phenomenon in Artificial Intelligence. *Syenza News*. 2024.
31. Haverinen J, Turpeinen M, Falkenbach P, Reponen J. Implementation of a new Digi-HTA process for digital health technologies in Finland. *International Journal of Technology Assessment in Health Care*. 2022;38(1):e68.
32. Haverinen J, Keränen N, Falkenbach P, Maijala A, Kolehmainen T, Reponen J. Digi-HTA: Health technology assessment framework for digital healthcare services. *Finnish Journal of eHealth and eWelfare*. 2019;11(4):326–41.
33. Rasmussen B S HMG, Fasterholdt I, Kidholm K, . Model for ASsessment of Artificial Intelligence Centre for Clinical Artificial Intelligence; [Available from: <https://caix.com/projects/previous-projects/mas-ai>].
34. Fasterholdt I, Naghavi-Behzad M, Rasmussen BSB, Kjølhede T, Skjøth MM, Hildebrandt MG, et al. Value assessment of artificial intelligence in medical imaging: a scoping review. *BMC Medical Imaging*. 2022;22(1):187.
35. EUnetHTA. HTA core model - Guiding principles on use 2015 [Available from: [https://www.eunethta.eu/wp-content/uploads/2018/01/The-HTA-Core-Model\\_Guiding-principles-on-use\\_20151218.pdf](https://www.eunethta.eu/wp-content/uploads/2018/01/The-HTA-Core-Model_Guiding-principles-on-use_20151218.pdf)].
36. Pfadenhauer LM, Gerhardus A, Mozygemba K, Lysdahl KB, Booth A, Hofmann B, et al. Making sense of complexity in context and implementation: the Context and Implementation of Complex Interventions (CICI) framework. *Implement Sci*. 2017;12(1):21.
37. Karargyris A, Umeton R, Sheller MJ, Aristizabal A, George J, Wuest A, et al. Federated benchmarking of medical artificial intelligence with MedPerf. *Nature Machine Intelligence*. 2023;5(7):799-810.
38. HTx. Next Generation Health Technology Assessment European Union; n.d. [Available from: <https://www.htx-h2020.eu/>].

39. Garcia-Saez G, Goettsch W, Driessen J H M, Nemeth B, Petrova G, Siirtola P, Röning J, Zempenyl A T, Hernando M E, . Next Generation Health Technology Assessment to support patient-oriented, societally oriented, real-time decision-making in Diabetes. The HTx Consortium; 2020.
40. Gyldmark M, Lampe K, Ruof J, Pöhlmann J, Hebborn A, Kristensen FB. IS THE EUNETHTA HTA CORE MODEL® FIT FOR PURPOSE? EVALUATION FROM AN INDUSTRY PERSPECTIVE. *International Journal of Technology Assessment in Health Care*. 2018;34(5):458-63.
41. Moser A, Korstjens I. Series: Practical guidance to qualitative research. Part 3: Sampling, data collection and analysis. *European Journal of General Practice*. 2018;24(1):9-18.
42. Turner C, Astin F. Grounded theory: what makes a grounded theory study? *Eur J Cardiovasc Nurs*. 2021;20(3):285-9.
43. Eppich WJ, Olmos-Vega FM, Watling CJ. Grounded Theory Methodology: Key Principles. In: Nestel D, Hui J, Kunkler K, Scerbo MW, Calhoun AW, editors. *Healthcare Simulation Research: A Practical Guide*. Cham: Springer International Publishing; 2019. p. 127-33.
44. Belgrave LL, Seide K. Grounded Theory Methodology: Principles and Practices. In: Liamputtong P, editor. *Handbook of Research Methods in Health Social Sciences*: Springer Singapore; 2019. p. 299-316.
45. van der Bijl-Brouwer M, Dorst K. Advancing the strategic impact of human-centred design. *Design Studies*. 2017;53:1-23.
46. van Gemert-Pijnen L, Kelders S M, Kip H, Sanderman R, . *eHealth Research, Theory and Development*. New York: Routledge; 2018.
47. Melles M, Albayrak A, Goossens R. Innovating health care: key characteristics of human-centered design. *International Journal for Quality in Health Care*. 2020;33(Supplement\_1):37-44.
48. van der Bij JD, Dijkstra L, de Vries G, Walburg J. Improvement and renewal of healthcare processes: results of an empirical research project. *Health Policy*. 1999;48(2):135-52.
49. Badreddin O. Thematic Review and Analysis of Grounded Theory Application in Software Engineering. *Advances in Software Engineering*. 2013;2013.
50. Scharwächter V. *Operationaliseren van Concepten*: Scribbr; 2022 [updated 17 oktober 2022]. Available from: <https://www.scribbr.nl/scriptie-structuur/stappenplan-operationaliseren/>.
51. *Simply Stakeholder Guide to Stakeholder Analysis*: Simply Stakeholders; 2023 [Available from: <https://simplystakeholders.com/stakeholder-analysis/>].
52. Van Haesendonck L, Ruof J, Desmet T, Van Dyck W, Simoens S, Huys I, et al. The role of stakeholder involvement in the evolving EU HTA process: Insights generated through the European Access Academy's multi-stakeholder pre-convention questionnaire. *J Mark Access Health Policy*. 2023;11(1):2217543.
53. Unsworth H, Wolfram V, Dillon B, Salmon M, Greaves F, Liu X, et al. Building an evidence standards framework for artificial intelligence-enabled digital health technologies. *The Lancet Digital Health*. 2022;4(4):e216-e7.
54. Gomez Rossi J, Feldberg B, Krois J, Schwendicke F. Evaluation of the clinical, technical, and financial aspects of cost-effectiveness analysis of artificial intelligence in medicine: scoping review and framework of analysis. *JMIR Medical Informatics*. 2022;10(8):e33703.
55. Hogervorst MA, Vreman RA, Mantel-Teeuwisse AK, Goettsch WG. Reported Challenges in Health Technology Assessment of Complex Health Technologies. *Value in Health*. 2022;25(6):992-1001.
56. Unsworth H, Dillon B, Collinson L, Powell H, Salmon M, Oladapo T, et al. The NICE Evidence Standards Framework for digital health and care technologies – Developing and maintaining an innovative evidence framework with global impact. *DIGITAL HEALTH*. 2021;7:20552076211018617.
57. Julian E, Gianfrate F, Sola-Morales O, Mol P, Bergmann J-F, Salmonson T, et al. How can a joint European health technology assessment provide an 'additional benefit' over the current standard of national assessments? *Health Economics Review*. 2022;12(1):30.
58. EUPATI. *Health Technology Assessment process: Fundamentals: EUPATI 2015* [updated 23 november 2015]. Available from: <https://toolbox.eupati.eu/resources/health-technology-assessment-process-fundamentals/>.

59. Adeoye-Olatunde OA, Olenik NL. Research and scholarly methods: Semi-structured interviews. JACCP: JOURNAL OF THE AMERICAN COLLEGE OF CLINICAL PHARMACY. 2021;4(10):1358-67.
60. the patients association The Digital Coalition n.d. [Available from: <https://www.patients-association.org.uk/Listing/Category/the-digital-coalition>].
61. Zorginstituut Nederland Over ons: Zorginstituut Nederland; n.d. [Available from: <https://www.zorginstituutnederland.nl/over-ons>].
62. Kang E, Hwang H-J. The importance of anonymity and confidentiality for conducting survey research. Journal of Research and Publication Ethics. 2023;4(1):1-7.
63. Alhojailan M I. Thematic Analysis: A critical review of its process and evaluation. 2012.
64. Tanujaya B, Prahmana RCI, Mumu J. Likert scale in social sciences research: Problems and difficulties. FWU Journal of Social Sciences. 2022;16(4):89-101.
65. Twisk J W R. Inleiding in de toegepaste biostatistiek. Fourth ed. Houten: Bohn Stafleu van Loghum; 2016.
66. Canva. Over Canva n.d. [Available from: <https://www.canva.com/>].
67. Haraldsen IH, Hatlestad-Hall C, Marra C, Renvall H, Maestú F, Acosta-Hernández J, et al. Intelligent digital tools for screening of brain connectivity and dementia risk estimation in people affected by mild cognitive impairment: the AI-Mind clinical study protocol. Front Neurobot. 2023;17:1289406.
68. DeMartino GF. The specter of irreparable ignorance: counterfactuals and causality in economics. Rev Evol Polit Econ. 2021;2(2):253-76.
69. Marka AW, Luitjens J, Gassert FT, Steinhelfer L, Burian E, Rübenthaler J, et al. Artificial intelligence support in MR imaging of incidental renal masses: an early health technology assessment. European Radiology. 2024.
70. Noble H, Smith J. Issues of validity and reliability in qualitative research. Evidence Based Nursing. 2015;18(2):34.
71. Lukkien DRM, Nap HH, Buimer HP, Peine A, Boon WPC, Ket JCF, et al. Toward Responsible Artificial Intelligence in Long-Term Care: A Scoping Review on Practical Approaches. The Gerontologist. 2021;63(1):155-68.
72. Merlin T, Street J, Carter D, Haji Ali Afzali H. Challenges in the Evaluation of Emerging Highly Specialised Technologies: Is There a Role for Living HTA? Appl Health Econ Health Policy. 2023;21(6):823-30.

## Appendix

A: Overzicht tot dusver gepubliceerde HTA(-gebaseerde) raamwerken voor medische AI

PROPERTY OF I. DE BRUIN (FOUNDER OF AISS)

<b>Bestaande HTA-gebaseerde raamwerk</b>	<b>Overkoepelend: omvat alle typen medische AI toepassingen?</b>	<b>Multidisciplinair: Juiste stakeholders / perspectieven betrokken?</b>	<b>Unbiased en transparant?</b>	<b>Bevat alle benodigde aspecten en criteria?</b>	<b>Ontwikkeling / totstandkoming?</b>
<b>Digi-HTA (31, 32)</b>	Vooral gericht op AI toepassingen voor monitoring buiten zorginstellingen, zoals mHealth (mobile health applications) (31, 32).	Het perspectief van de 'decision-makers' is niet meegenomen in de ontwikkeling, waardoor ziekenhuizen geen gebruik maken van de Digi-HTA (31).	Afgeleverde informatie door de makers en andere betrokken stakeholders wordt gecontroleerd door systematische review artikelen, HTA experts en cyber security experts (31).	Het bevat alleen effectiviteit, veiligheid, kosten, technische karakteristieken en patiënt- en organisatorische overwegingen. Sociale, ethische en juridische aspecten en criteria zijn niet meegenomen (31).	Op basis van de 'Hyteairo' en 'KATI' programma's van de Finse overheid en criteria uit de 'Cyber Health Project' (31).
<b>MAS-AI (Model for Assessing the value of Artificial Intelligence) (33, 34)</b>	In eerste instantie voornamelijk gericht op medische beeldvorming en op het zorgstelsel van Denemarken(34). Daarna wordt het toepassingsgebied steeds verder uitgebreid (33, 34).	Het is ontwikkeld door meerdere disciplines uit Denemarken, namelijk vooraanstaande onderzoekers, 'decision-makers', patiënten verenigingen en andere onderzoekers (34). Het is onduidelijk of zorgverleners, de overheid en AI ontwikkelaars hierin meegenomen zijn (33, 34).	Dit is onduidelijk te achterhalen (33, 34).	Specifieke klinische outcomes voor het evalueren van AI technologieën zijn erkend als relevant, maar zijn niet meegenomen (34). De 'early MAS-AI' bestaat uit vier evaluatiegebieden (33): "The health problem and the current se of technology", "Technology", "Ethical aspects" en "Legal aspects". Daarna wordt een 'full MAS-AI' gedaan die bestaat uit vijf andere evaluatiegebieden (33): "Safety", "Clinical aspects", "Economy", "Organisation" en "Patient perspectives".	Begon met een scoping review, daarna interviews met zes vooraanstaande onderzoekers uit Denemarken, en als laatste twee workshops met decision-makers, patiënten verenigingen en onderzoekers (33, 34).
<b>HTA Core Model (35, 40)</b>	Toepasbaar voor alle soorten zorgtechnologieën (35, 40)	Het is onduidelijk welke disciplines betrokken waren bij het ontwerpen applicatieproces (35, 40)	De input en gebruikte methoden zijn niet altijd transparant weergegeven (40).	Het raamwerk kan in meerdere zorginstellingen in Europa gebruikt worden (40). De time-horizons, tijdlijnen, uitvoering en transparantie van het raamwerk moeten verbeterd worden en meer gestandaardiseerd met HTA's in andere zorgsettingen (40)	Het is onduidelijk hoe het raamwerk is ontwikkeld en door wie (35, 40)
<b>Living HTA (20)</b>	Het is niet breder toepasbaar dan zeer gespecialiseerde zorgtechnologie, zorgpaden en voor het ontdekken van 'unmet needs' vanuit het patiënt en klinisch perspectief. Het is niet duidelijk of het voor alle medische AI toepassingen gebruikt kan worden (72).	Het is onduidelijk of de 'Living HTA' is ontwikkeld op basis van literatuuronderzoek en/of praktijkonderzoek (zoals interviews) (20, 72).	Triggerpoints voor updates worden bepaald door feedback van stakeholders en wanneer nieuw bewijs naar voren komt (20).	Bij grote veranderingen in het bewijs en/of data moet het raamwerk hoogstwaarschijnlijk aangepast worden en nieuwe 'outcomes' kunnen toegevoegd worden (20). Updates kunnen alleen selectief uitgevoerd worden en niet bij alle onderdelen van het HTA raamwerk (20).	Gebaseerd op hoe 'Living systematic reviews and meta-analyses' en updates van standaard HTA raamwerken worden uitgevoerd (20, 72).

<b>CICI framework (Context and Implementation of Complex Interventions) (36)</b>	Het is toepasbaar bij complexe zorg-technologieën in de Europese zorgsector (36).	Het is onduidelijk welke disciplines betrokken waren bij het ontwerp- en applicatieproces (36).	Het is namens de Europese Unie (EU) opgezet en gefinancierd (36). De inhoud van het raamwerk is meerdere keren geëvalueerd door theoretische methoden (zoals scoping reviews) en praktijk methoden ('Rapid assessments'), en het raamwerk is extern gereviewd (36).	Context en implementatie aspecten en criteria worden expliciet meegenomen (36). Het lijkt geen opzichzelfstaande raamwerk te zijn, maar als toevoeging van een standaard HTA raamwerk (36).	Het raamwerk is opgesteld door eerst een scoping review te doen, daarna te testen in 'rapid assessments', vervolgens is er een 'pragmatic utility concept analysis' uitgevoerd en als afsluiting van het onderzoek aangepast en getest met bepaalde systematische reviews (36).
<b>MedPerf (37)</b>	Alleen gericht op data om medische AI modellen te evalueren en te benchmarken door deze modellen te testen met testdata en de ICT infrastructuur te testen (37).	Gezamenlijk ontwikkeld door de FDA (Food and Drug Administration), de EMA (European Medicines Agency) en de CDSCO in India (Central Drugs Standard Control Organisation) (37). Daarnaast zijn 'data owners' (de wettelijke eigenaren van data) ook betrokken bij het ontwerpproces (37).	Een 'Benchmark committee' controleert de AI modellen en de testdata (37).	Het kijkt alleen naar technische algemene aspecten van de achterliggende modellen van een medische AI toepassing, dus niet de bijbehorende algoritmes en de medische AI toepassing als geheel (37).	Er is gebruik gemaakt van 'federated evaluation' om makkelijk en snel de AI modellen te kunnen delen (37).
<b>HTx (Next Generation Health Technology Assessment) (38)</b>	Gericht op medische AI toepassingen voor EPD-ondersteuning (Elektronisch Patiënt Dossier) en 'real-time decision-making' (38).	Het is onduidelijk vermeldt of het HTx raamwerk met patiënten en zorgverleners is ontworpen (38).	Alle betrokken stakeholders, HTA instanties en de betalende partijen kunnen het raamwerk en de informatie in dit raamwerk inzien (38) [34]. De input komt van RWD (Real World Data) en RCT studies (Randomised Controlled Trials) (39).	Het is bedoeld voor lidstaten van de Europese Unie en kan ook wereldwijd toegepast worden (38). Vooral patiënt- en maatschappelijk geïntereerd (39).	Vier verschillende case studies zijn uitgevoerd om de toepasselijkheid bij verschillende aandoeningen, technologieën, lifestyle interventies en behandelmethoden (39).

## B: Literatuuronderzoek opzet

Om een theoretische basis te leggen werd er eerst een (korte) literatuuronderzoek uitgevoerd om systematisch relevante literatuur te verzamelen naar de relevante aspecten en criteria voor zowel het klinisch en economisch evalueren van en bestaande HTA raamwerken voor medische AI toepassingen in zijn algemeenheid. Dit is gedaan vanaf 10 april 2024 tot en met 10 mei 2024 door de onderzoeker zelf met eventuele input van de eerste begeleider. De zoektermen waren gebaseerd op de variabelen en onderwerpen in de centrale onderzoeksvraag en deelvragen van dit onderzoek. Deze zijn:

“(“Health Technology Assessment”) OR (HTA) OR (“HTA-based framework”))” AND “(“Artificial Intelligence”) OR (AI)” AND “((criteria) OR (“necessary data”))” AND “(“stakeholder involvement”))” AND “((develop\*) OR (creat\*))”. De gebruikte databasen waren Google Scholar, Pubmed, Scopus/Elsevier en Springer. Hierbij wordt gefilterd op artikelen vanaf 2014, omdat ontwikkelingen van medische AI toepassingen voor verschillende applicatiegebieden in de zorgsector vanaf dat jaar wereldwijd een exponentiële groei doormaakten (3). Hierbij is ook gebruik gemaakt van de ‘sneeuwbalmethode, waarbij maximaal drie keer is doorgeklikt in de referenties van een relevant artikel/onderzoek naar een andere relevante artikel/onderzoek om informatie te vergaren (40, 45). De relevantie van de gevonden (wetenschappelijke) artikelen/onderzoeken is beoordeeld aan de hand van de variabelen en onderwerpen van de centrale onderzoeksvraag en deelvragen (4). De resultaten werden verwerkt in een operationalisatieschema (appendix A), omdat hierdoor op een systematische manier abstracte begrippen (concepten) gedefinieerd en meetbaar gemaakt werden (operationaliseren) door ze te ontleden in dimensies met eventueel bijbehorende subdimensies. Deze (sub)dimensies zijn opgesplitst in variabelen met indicatoren, in het geval van kwantitatieve data, om als input te dienen voor de survey, eerste aanzet van de HTA-mAix en het identificeren van lastig waarneembare fenomenen gedurende de analyses van de onderzoeken (4, 48, 49). Tevens diende de resultaten als input voor het theoretisch kader en het interpreteren van de uitkomsten van de interviews en survey in de discussie.

### C: Operationalisatieschema

\*Not included in online document.

### D: Inclusie- en exclusiecriteria en stakeholderanalyse

<b>Categorie</b>	<b>Inclusiecriteria</b>	<b>Exclusiecriteria</b>
<i>Perspectief</i>	Medisch en klinisch (52-55)	Farmaceutisch (51, 54)
	Economisch en toezichhoudend (37, 51, 52)	Niet-zorg gerelateerd [51-54]
	Patiënt (37, 51, 53, 54)	
	Academisch (54)	
	Technologisch (37, 51, 54)	
	Organisatorisch (37, 51, 54, 55)	
<i>Kennis en expertise</i>	Ontwikkelingen van AI toepassingen in de zorgsector (56)	Ontwikkelen en/of toepassen van HTA raamwerk(en) voor de farmaceutische industrie en/of andere medicamenteuze behandeling waarbij AI geen rol speelt (57)
	Ervaring met evaluatie, gebruik en/of onderzoek naar medische AI toepassingen	HTA raamwerk(en) toepassen en/of ontwikkelen voor ziekte preventieprogramma's waarbij medische AI toepassingen geen rol spelen (57)
	Toepassen van HTA raamwerken bij complexe zorginnovaties, met name AI-gerelateerd (54, 56)	

Implementatie van complexe zorginnovaties, met name AI-gerelateerd (57)	
Ervaring met het evalueren van medische AI toepassingen met HTA raamwerk(en) en/of gezondheid economische evaluaties (54, 56)	
Toepassen en/of ontwikkelen van evaluatiecriteria voor AI toepassingen in de zorgsector of één of meerdere zorgapplicatiegebieden (zoals diagnostiek) (20, 56)	
Onafhankelijk (57)	

### Stakeholderanalyse – “power/interest-grid”

	Weinig belang (interest)	Veel belang
Weinig macht (power)		<ul style="list-style-type: none"> <li>- (Nederlandse) onderzoekers op het gebied van HTA raamwerken en/of gezondheid economische evaluaties voor een of meerdere medische AI toepassingen;</li> </ul>
Veel macht		<ul style="list-style-type: none"> <li>- Zorginstituut Nederland (58);</li> <li>- The Patient Association – the digital coalition for AI, Data and Digital technology in health (59);</li> <li>- Microsoft + Nuance communications</li> </ul>

### E: Interviewschema

#### Interviewschema (interview guide) bachelor thesis HTA-raamwerk voor AI

Interviewer (interviewer):

Interviewee (geïnterviewde):

Date (datum):

Location (locatie):

Start time (start tijd):

End time (eind tijd):

---

## Introduction (introdactie)

- Bedanken voor deelname aan het interview.
- Mijzelf voorstellen, rationale en doel van mijn onderzoek en dit interview uitleggen (**vb:**
- Reden voor dit interview met desbetreffende stakeholder
- Verifiëren of getekende IC (Informed Consent) nog klopt. (anoniem blijven, informatie gebruiken voor thesis en interview opnemen)
- Aangeven recht om vragen niet te beantwoorden.
- Uitleggen hoe interview gaat verlopen: eerst wat achtergrond vragen stellen, exploratief onderzoek waardoor brede vragen gesteld worden, daarna meer in op de criteria en aspecten voor het evalueren van AI
- Het interview zal ongeveer 30 tot 60 minuten duren.
- Heeft u nog vragen vooraf?
- Opname starten.

## Interviewschema

### 1. Achtergrond informatie

#### Vertegenwoordiging

- Vragen de geïnterviewde zichzelf (kort) voor te stellen.
- Vragen voor een uitleg (omschrijving) over zijn/haar functie.
- Land en stakeholdergroep vertegenwoordiging

#### Ervaring

- Heeft u ervaring met het evalueren van medische Artificial Intelligence toepassingen voor in de gezondheidszorg?
- **Indien van toepassing:** Heeft u ervaring met het uitvoeren of ontwikkelen van HTA-raamwerken voor het evalueren van medische AI-toepassingen?
- Maakt u gebruik (of heeft u een keer gebruik gemaakt) van HTA raamwerken/ een HTA-raamwerk in uw functie/werk?

### 2. Benodigde input/output

### Benodigde perspectieven

- Vanuit welke perspectieven moeten medische AI-toepassingen geëvalueerd worden om een juist beeld te krijgen van o.a. de (verwachte) voordelen, (verwachte) effecten en (verwachte) behaalde winst van deze technologieën, volgens u?
- Hoe zou men het gebruikersperspectief moeten meenemen tijdens het evalueren/ uitvoeren van een HTA voor medische AI-toepassingen, volgens u?
- Welke directe en indirecte consequenties zijn van toepassing bij medische AI-toepassingen, volgens u?

Welke informatie/data zijn nodig om dit te bepalen/te schatten?

### (Klinische) effectiviteit

- Welke informatie/data zijn van belang voor het evalueren van de klinische effectiviteit van medische AI-toepassingen in zijn algemeenheid?

Hoe zou dit verzameld/inzichtelijk gemaakt moeten worden?

- Welke health outcomes zijn van belang voor het evalueren van medische AI-toepassingen?

Zijn verschillende criteria van belang afhankelijk van het type medische AI-toepassing of per zorggebied, volgens u?

Zouden verschillende health outcomes verzameld moeten worden afhankelijk van het perspectief (zoals patiënt perspectief, zorgverlener perspectief), volgens u?

### Benodigde technische informatie van medische AI

- Welke informatie/data over een medische AI-toepassingen of medische AI-toepassingen in zijn algemeenheid is van belang voor het zo volledig mogelijk evalueren van deze technologie(ën)?

Hoe zou deze informatie/data verzameld/inzichtelijk gemaakt moeten worden?

Wie is hiervoor verantwoordelijk?

### Kosteneffectiviteit

- Welke informatie/data zijn van belang voor het bepalen/inzichtelijk maken van de kosteneffectiviteit van medische AI-toepassingen?

Hoe zou deze informatie/data verzameld/inzichtelijk gemaakt moeten worden?

Wie zouden hiervoor verantwoordelijk moeten zijn?

- Hoe zouden de verwachte impacts (economische, financiële kant) bepaald/geschat moeten worden voor medische AI-toepassingen in zijn algemeenheid?
- Wie moeten hiervoor verantwoordelijk zijn?

Welke informatie/data zouden hiervoor nodig moeten zijn?

- Hoe moet de kosteneffectiviteit van medische AI-toepassingen (in zijn algemeenheid) geschat worden, volgens u?

Zijn standard CEA's hiervoor een goede manier, volgens u? (zo ja: verder toelichten, zo nee: welke formules zijn hier wel geschikt voor?)

Is dit verschillend per medische AI toepassing/zorggebied?

### 3. Randvoorwaarden

#### Her-evaluatie (re-evaluation)

- Denkt u dat het nodig is om een medische AI-toepassing te her-evalueren omdat het zich blijft ontwikkelen?

Wanneer zou de evaluatie (en her-evaluaties) uitgevoerd moeten worden?

Door wie zou dit gedaan moeten worden (wie is verantwoordelijk voor de uitvoering), volgens u?

- Hoe zouden medische AI-toepassingen het beste geëvalueerd (en/of her-evalueert) moeten worden, volgens u?

#### Incrementeel evalueren

- Denkt u dat het nodig is om medische AI-toepassingen incrementeel te evalueren aangezien het zich steeds verder aan het ontwikkelen is?

#### Wet- en regelgeving

- Welke nationale en internationale wet- en regelgeving zijn van belang bij het ontwerpen en/of uitvoeren van een universeel toepasbare HTA gebaseerde raamwerk voor het evalueren van medische AI-toepassingen?

### Stakeholder betrokkenheid

- Hoe moet de samenwerking tussen de betrokken disciplines eruit komen te zien tijdens de uitvoering/toepassen van dit HTA-gebaseerd raamwerk?

### **Afsluiting**

**Afsluitende vraag1:** denkt u, vanuit uw perspectief, dat het mogelijk is om een raamwerk te maken voor het volledig (holistisch) evalueren van medische AI technologieën/toepassingen in zijn algemeenheid (of kan dit alleen per zorggebied)?

Zou u dit verder kunnen toelichten?

**Afsluitende vraag 2:** Denkt u, vanuit uw perspectief, dat het in de praktijk mogelijk is om een universeel toepasbare HTA-raamwerk te ontwerpen voor het incrementele assessment van medische AI-applicaties die zowel algemeen als per specifieke applicatiegebied toegepast kan worden?

Zou u dit verder kunnen toelichten?

### Afsluitende informatie

- Heeft u nog vragen of opmerkingen, aanvullingen?
- Nogmaals bedanken voor deelname aan het interview voor mijn bachelor afstudeerscriptie.
- Nog een keer toelichten hoe de informatie uit het interview geanalyseerd en verwerkt gaat worden.
- Vragen of desbetreffende stakeholder open staat om mijn online survey (1) te willen invullen en te delen met collega's.
- Opname stoppen.
- **Aangepast o.b.v. Informed Consent formulier:** Aangeven dat geïnterviewde een copy van het geschreven onderzoek ontvangt waarin zijn/haar informatie is gebruikt, wanneer bachelor thesis is afgerond.

### [F: Informed Consent interviews](#)

**Informed Consent for 'first attempt to design a universally applicable HTA-based framework for the (incremental) assessment of medical Artificial Intelligence (AI) applications: an exploratory mixed-methods approach'**

After completing and signing the Informed Consent (this includes "Information sheet" and "Informed Consent form") you will receive a digital copy.

### **Information sheet**

Research project title: First attempt to design a universally applicable HTA-based framework for the (incremental) assessment of medical Artificial Intelligence (AI) applications: an exploratory mixed-methods approach

Dear .....,

Thank you for agreeing to be interviewed for this research project. This letter explains what the above-mentioned research project entails and how the interviews for this research project will be conducted. Please take time to read the following information carefully. If any information is not clear kindly ask questions using the contact details of the researchers provided at the end of this letter.

### **Purpose of the research project**

This research project aims to facilitate decision-makers, healthcare professionals, healthcare institutions, regulatory bodies, etcetera with the necessary information to appropriately assess the value of medical Artificial Intelligence (AI) applications. Therefore, this research project will make a first try at designing a new universally applicable Health Technology Assessment (HTA) based framework for the (incremental) assessment of medical AI applications both in specific care-area's and in general.

### **Benefits and risks of participating in this research project**

Benefits of participation are the contribution to a universally applicable HTA-based framework for the (incremental) assessment of medical AI applications to improve healthcare systems worldwide. Furthermore, your expertise will help understand what the issues and missing information are surrounding the integration of medical AI applications in healthcare systems. Lastly, you will contribute to improving the collaboration between many stakeholders in healthcare, help to break down barriers and transparency needed to improve the understanding of the benefits that AI technologies can bring to healthcare systems worldwide.

Risks of participation are little to none, there is only the risk of the loss of privacy.

This research project has been reviewed and approved

### **Procedure for withdrawal from the research project**

As mentioned before, we don't anticipate that there are any risks associated with your participation. However, you have the right to stop the interview or withdraw from this research project at any time. Withdrawal can be done by sending an email to me, the researcher of this research project (see contact details below), either before the planned interview or after the interview. During the interview, you have the right to discontinue the interview.

### **Participation involvement**

- Participation is voluntary.
- You have the right to refuse to answer a question without providing a reason.
- Duration of participation in the interview will take place once on an agreed upon day and will be approximately a minimum of 30 minutes and a maximum of 60 minutes.

- The personal information that will be collected before and during the interview is your name, the stakeholder group you represent, experiences with HTA frameworks, experiences with medical AI applications and the country you represent.

### **Information usage**

The information collected from the interview will be transcribed by the researcher and analyzed to identify needed aspects and criteria for the design of the universally applicable HTA-based framework for the (incremental) assessment of medical AI applications. Other information will be analyzed and processed to explain the context behind the design of the newly created HTA-based framework. The information that will be used for this research project will be processed anonymously by me, the researcher of this research project. This will be done by removing your name and the organization you work for from the interview transcript and deleting the recording of the interview after finishing the research project. This indicates that the information cannot be traced back to you.

Due to the regulations of the University of Twente (UTwente) the research project will be published on [essay.utwente.nl](http://essay.utwente.nl) when finished. To guarantee the used information will still be anonymous when published on this website, the appendixes containing the transcribed interviews will not be accessible for visitors of this website.

### **Contact details**

Researcher: Irene de Bruin,

Supervisor 1: .....,

Supervisor 2: .....,

The BMS Ethics Committee/domain Humanities & Social Sciences:

### **Contact Information for Questions about Your Rights as a Research Participant**

If you have questions about your rights as a research participant, or wish to obtain information, ask questions, or discuss any concerns about this study with someone other than the researcher(s), please contact ....

### **Informed Consent form**

Research project title: First attempt to design a universally applicable HTA-based framework for the (incremental) assessment of medical Artificial Intelligence (AI) applications: an exploratory mixed-methods approach

Researcher: Irene de Bruin

Ethical procedures for academic research undertaken by institutions require that interviewees explicitly agree to being interviewed and how the information contained in your interview will be used. This Informed Consent form is necessary that you understand the purpose of your involvement and that you agree to the conditions of your participation.

### **Taking part in the research project**

After you have read the accompanying information from the information sheet (see above), would you sign this form to certify that you approve the following:

- The interview will be recorded and a transcript will be produced from the recorded interview.

- The transcript of the interview will be analyzed by the researcher and possibly the involved supervisors.
- Access to the interview transcript will be limited to the researcher, possibly the involved supervisors and academic colleagues with whom might be collaborated as part of the research process.
- After completing this research project, the actual recording of the interview will be deleted from the device used and the transcript of the interview will be kept for researchers copy only.
- Any variation of the conditions above will only occur with your further explicit approval.

Name:	Date:	Signature:

### Quotation agreement

With regard to being quoted, please tick the appropriate boxes next to the statements.

Statement	Yes	No
I wish to receive the transcript of the interview to be given the opportunity to correct any factual errors.		
I agree to be quoted directly.		
I wish to remain anonymous when the information is analyzed and processed in the research project.		
I wish to receive a copy of the results pertaining to my participation in this research project after the research project is finished.		

**Study contact details for further information about the Informed Consent form: Irene de Bruin,.....**

G: Online survey

**\*Not included in online document.**

H: Resultaten thematische analyses

Aspect	Criteria
Context	Beschrijving van de algemene doeleinden van een medische AI-toepassing
	Mandaat (digitaal toegangsgebied) van de AI algoritme/ AI model
	Beschrijving van het beoogde gebruik (gebruikersaanwijzingen)
	Beschrijving van de benodigde kwalificaties, kennis en vaardigheden om de medische AI-toepassing te gebruiken

	Consistentie in de prestaties
	Onderzoeksvraag volgens PICOTP (Patient/Population, Intervention/Innovation, Comparison, Outcomes, Time horizon and Perspective(s) or setting))
	Multidisciplinair proces
	Gespecificeerde primaire en secundaire (gezondheid) uitkomsten behorende bij specifieke applicatie gebied en per type medische AI-toepassing
	Gepaste alternatieven (comparators)
	Beschrijving van de beoogde eindgebruikers
	Beschrijving van het bewijs en data dat is gebruikt voor de ontwikkeling van de medische AI-toepassing
	Evaluëren van de organisatorische geschiktheid
	Gebruikersgeschiktheid evalueren
	Geschiktheid wet- en regelgeving evalueren
	Evaluëren van de betrouwbaarheid
	De potentiële leercurve door het gebruik van de medische AI-toepassing
	Identificeren welke stakeholders voordelen zullen ervaren en welke stakeholders nadelen zullen ervaren door het gebruik van medische AI-toepassing
<b>Betrokken perspectieven</b>	Medisch, klinisch perspectief
	Patiënt, client perspectief
	Evaluator, toetsend perspectief
	AI-ontwikkelaars perspectief
	(Inter)nationale overheidsinstanties, wet- en regelgeving perspectief
	Zorgverzekeraar perspectief
<b>(Klinische) effectiviteit</b>	Klinische toepasbaarheid (e.g. is the medical AI application applicable in practice)
	Impact op (klinische) workflow
	Eindgebruikers acceptatie van de medische AI-toepassing bepalen
	Impact op (klinische) werklust en werkuren
	Impact op toegankelijkheid van zorg
	Impact op kwaliteit van leven (QoL) van de betrokken stakeholders
	Impact op baantevredenheid scores van zorgverleners en/of andere direct betrokken stakeholders
	Impact op patiënttevredenheid scores
	Bepalen van (klinische) effectiviteit drempelwaarden
	Kans op falen
	Identificeren van andere voor- en nadelen
	Identificeren van alle mogelijke gezondheidsrisico's van de medische AI-toepassing voor de eindgebruikers andere direct betrokken stakeholders
	Kwalitatief bewijs
	Subjectief bewijs
<b>Kosteneffectiviteit (gezondheid)</b>	Zo veel mogelijk includeren en kwantificeren van de effectiviteitscriteria uit het aspect (klinische) effectiviteit

<b>economische evaluaties)</b>	
	Juiste gezondheid economische evaluaties gebruiken
	Juiste tijdshorizon
	Mogelijkheid om de medische AI-toepassing uit te breiden
	Impact op huidige en toekomstige zorglasten
	Directe en indirecte (verwachte) consequenties op zorguitgaven voor desbetreffende zorginstelling (e.g. hospital, nursing home)
	Budget-impact van het gebruik van de medische AI-toepassing (voor de zorginstelling en nationale zorguitgaven)
	Impact op de betaalbaarheid van zorg
	Impact op arbeidskosten
	Investeringskosten voor het gebruik van de medische AI-toepassing
	Alle kosten, inclusief verzekeringskosten, van de beoogde applicatie gebied
	Het vastgestelde tarief voor het bekostigen van de medische AI-applicatie
	Bepalen van nieuwe Willingness To Pay (WTP) drempelwaarde(n) voor medische AI-toepassingen (voor zorgverzekeraars, overheidsinstanties en betalende partijen)
	Kosten voor data opslag
	Identificeren van potentiële subsidies voor de medische AI-toepassing
	Kosten voor het runnen van de AI algoritme(s)/ AI model(len)
<b>Maatschappelijke impact</b>	Impact op kwaliteit van zorg op een nationaal niveau
	Impact van de medische AI-toepassing op de socio-economische status van de betrokken patiëntenpopulatie(s) en andere betrokken stakeholders
	Indien nodig, (zoals e-Health AI), impact en toepassingsmogelijkheden bij digibete of niet-digitaal vaardige patiëntenpopulaties wanneer de medische AI-toepassing "fysieke" zorgpaden vervangt
	Milieu impact van de medische AI-toepassing
	(Inter)nationaal vaststellen van "goed gebruik van medische AI-toepassingen" criteria en procedures om de bruikbaarheid te kunnen evalueren
	Directe en indirecte ethische consequenties voor eindgebruikers en andere betrokken stakeholders identificeren
<b>(Technische) algemene informatie van een medische AI-toepassing</b>	Uitlegbaarheid (explainability: mechanisms or operating systems of the medical AI application need to be understandable and clear for the end users)
	De gebruikte factoren of parameters in de AI algoritme(s)/ AI model(len)
	Transparantie over alle informatie met betrekking tot de AI algoritme(s)/ AI model(len)
	AI-algoritme(s) en/of AI-model(len) moeten deelbaar zijn met vergelijkbare zorginstellingen
	Beschrijving van de representativiteit van de gebruikte trainingsdatasets van de medische AI-toepassing

	Beschrijving van de onderlinge afhankelijkheden tussen databases die de medische AI-toepassing nodig heeft om te functioneren
	Beschrijving van de betrokkenheid van de eindgebruikers tijdens de ontwikkeling van de medische AI-toepassing
	Benodigde implementatie strategie bij specifieke applicatie gebied en per type medische AI-toepassing voor een succesvolle implementatie
	Monitoringsproces voor de medische AI-toepassing om de voortgang, effectiviteit en accuraatheid te evalueren
	Leercurve van de medische AI-toepassing
	Evalueren/beoordelen van nieuwe updates of releases
	Beschrijving van de gebruikte trainingsmethodes voor de medische AI-toepassing (zoals personal health train, Federated learning, Privacy Preserved learning)
	Uitbreidings- en/of samenvoegingsmogelijkheden van de AI-algoritmes en AI-modellen met andere systemen
	Vaststellen van alle mogelijke data veiligheidsrisico's van de medische AI-toepassing met behulp van risicoanalyses, inclusief risicoscores per geïdentificeerde risico
	Vaststellen van alle mogelijke veiligheidsrisico's van de medische AI-toepassing met behulp van risicoanalyses en veiligheidsassessments, inclusief risicoscores per geïdentificeerde risico
	Directe en indirecte (potentiële) cybersecurity consequenties bij gebruik van de medische AI toepassing

#### I: Kwantitatieve en kwalitatieve surveyresultaten

**\*Not included in online document.**

#### J: Resultaten online survey

**\*Not included in online document.**

*Codeboek: scores en interpretatie bij 5-puntsschaal vragen*

Score	Interpretatie
1	Not at all important: helemaal niet relevant
2	Slightly not important: niet heel erg belangrijk
3	Neutral: neutraal, weet ik niet/geen mening over
4	Slightly important: beetje belangrijk
5	Very important: zeer belangrijk

*Codeboek: respondentnummer met bijbehorende stakeholdergroep(en)*

**\*Not included in online document.**