

Experimenteel onderzoek AI-assistent seksuele gezondheid SAM

Onderzoeksresultaten en aanbevelingen voor een condoominterventie

November 2025

Soa Aids Nederland
Amsterdam

Inhoudsopgave

Samenvatting, conclusies en aanbevelingen	3
1 Introductie	8
1.1 Achtergrond	8
1.2 Aanpak ontwikkeling condoominterventie	8
2 Samenvatting resultaten co-creatie met de doelgroep	9
2.1 Achtergrond uitgangspunten condoominterventie	9
2.2 Belangrijkste resultaten co-creatiesessies met jongeren	10
2.2.1 Belangrijkste bevindingen	10
2.2.3 Belangrijkste aanbevelingen uit de co-creatiesessies	11
2.3 Vervolgstappen	11
3 Methode	13
3.1 Doel van het experimentele onderzoek	13
3.2 Onderzoeksdesign	13
3.3 Respondentselectie	14
3.4 Uitvoering onderzoek	15
3.5 Inhoud vragenlijst en dataverzameling	15
4 Resultaten	17
4.1 Respons	17
4.2 Beoordeling effectiviteit AI-assistent SAM	18
4.2.1 Inleiding	18
4.2.2 Kwaliteit van de dialogen en respondentselectie	18
4.2.3 Scores op determinanten	19
4.2.4 Conclusies effectiviteit condoominterventie SAM	24
4.3 Beoordeling usability chatbots	25
4.3.1 Inleiding	25
4.3.2 Beoordeling interactie	25
4.3.3 Beoordeling gebruiksvriendelijkheid	27
4.3.4 Beoordeling antwoorden	28
4.3.5 Beoordeling antwoorden	29
4.3.5 Sterke punten en suggesties voor verbetering volgens gebruikers	30
4.3.5.1 Wat vinden gebruikers goed aan SAM?	30
4.3.5.2 Wat zouden gebruikers verbeteren aan SAM	31
4.3.6 Conclusies usability chatbots	31

Samenvatting, conclusies en aanbevelingen

In opdracht van het ministerie van Volksgezondheid, Welzijn en Sport (VWS) heeft Soa Aids Nederland in de periode oktober 2024 – juli 2025 een innovatieve condoominterventie ontwikkeld gericht op de doelgroep heteroseksuele jongeren van 12 tot 25 jaar. De aanleiding hiervoor is de zorgwekkende stijging van het aantal gonorroe-infecties en het dalende condoomgebruik onder deze doelgroep in Nederland.

Condoomgebruik blijkt in alle typen relaties inconsistent, zowel bij eenmalige seksuele contacten als bij losse partners en vaste relaties. Onderzoek toont aan dat gewoontevorming een belangrijke rol speelt: wanneer partners in het begin van een (seksuele)relatie geen condoom gebruiken, is de kans klein dat dit gedrag later in diezelfde relatie verandert. Omdat jongeren steeds vaker seksuele contacten hebben buiten een vaste relatie, is gekozen voor de ontwikkeling van een taakgerichte AI-assistent, genaamd SAM. Deze ondersteunt jongeren in de beginfase van een nieuwe (seksuele) relatie en verwijst hen tegelijkertijd door naar betrouwbare informatie over seksuele gezondheid, seksueel overdraagbare aandoeningen (soa's) en condoomgebruik. Door deze aanpak blijft de interventie relevant voor alle jongeren in beginnende (seksuele) relatievormen. Vanwege beperkingen in tijd en budget richt de eerste versie van SAM zich op heteroseksuele jongeren en een beperkt aantal determinanten van condoomgebruik. Als SAM goed functioneert, kunnen in de toekomst zowel doelgroepen als determinanten worden uitgebreid.

Ontwikkeling en co-creatie

De interventie is ontwikkeld volgens de methode Intervention Mapping, waarbinnen een uitgebreide gedragsanalyse en co-creatie met de doelgroep belangrijke uitgangspunten zijn voordat de interventie wordt ontwikkeld. Uit 18 sessies met in totaal 20 jongeren (16 individuele gesprekken en twee duo-gesprekken) bleek een duidelijke voorkeur voor een AI-assistent (chatbot) als vorm vanwege het eigentijdse karakter, de laagdrempeligheid en anonimiteit. Jongeren gaven aan dat een voor hen interessante AI-assistent niet alleen over gezondheidsonderwerpen zoals soa's en condooms moet gaan, maar juist ook over lichtere thema's zoals daten, relaties en sekstechnieken. Op basis hiervan is SAM ontwikkeld als een taakgerichte chatbot. In tegenstelling tot open AI-modellen (zoals ChatGPT), garandeert een taakgerichte aanpak maximale betrouwbaarheid van de gegeven gezondheidsinformatie (door uitsluitend betrouwbare informatiebronnen te selecteren), een essentiële voorwaarde voor medische informatie en gezondheidsadviezen, zeker op het gebied van seksuele gezondheid.

Onderzoeksopzet

Om de effectiviteit en potentie van SAM te onderzoeken is een experimenteel onderzoek uitgevoerd onder 300 jongeren (N=150 in de experimentele groep met SAM, N=150 in de controlegroep). De effectiviteit is gemeten aan de hand van de geselecteerde determinanten van condoomgebruik (attitude, sociale norm, risicoperceptie en eigen-effectiviteit). De usability (waardering en acceptatie) van SAM is vergeleken met een andere AI-chatbot die nog in ontwikkeling is bij Soa Aids Nederland: Soatestwijzer.nl. Deze chatbot ondersteunt gebruikers van alle leeftijden en seksuele oriëntaties bij HIV- en soa-testen en hanteert een andere aanpak dan SAM. De vergelijking dient als referentiekader, niet om een keuze tussen beide AI-assistenten te maken.

Resultaten

Uit het onderzoek naar de *usability* blijkt dat gebruikers SAM ervaren als een vriendelijke, betrouwbare, gebruiksvriendelijke en veilige gesprekspartner. Op deze kenmerken scoort SAM op gelijk niveau als de AI-assistent van Soatestwijzer.nl. De bronvermeldingen en de objectieve toon zorgen voor een hoog vertrouwen in de informatie. Een belangrijk pluspunt dat vaak genoemd wordt, is dat SAM standaard doorverwijst naar betrouwbare bronnen en links toevoegt. Echter, in vergelijking met de open AI-chatbot van de controlegroep, wordt de interactie met SAM als minder vloeiend en meer ‘robotachtig’ ervaren. Andere verbeterpunten zijn:

- ✓ Variatie en relevantie: antwoorden en vervolgvragen zijn soms te standaard of niet passend.
- ✓ Reactiesnelheid: is voldoende maar mag sneller.
- ✓ Het persoonlijk herhalingsgebruik van SAM en de bereidheid om deze aan anderen aan te bevelen scoren gemiddeld net voldoende, maar blijven achter bij de gewenste niveaus. Aangezien de AI-assistent van Soatestwijzer.nl op deze aspecten beter presteert, ligt het voor de hand dat factoren zoals de nog ‘robotachtige’ persoonlijkheid van SAM en beperkte dialoogvaardigheden een rol spelen bij deze lagere waardering.

Uit de *usability* scores blijkt dat SAM de dialoog veel sterker stuurt naar onderwerpen zoals soa’s en condooms dan de AI-assistent van Soatestwijzer.nl. Dit verschil is logisch gezien de opdracht die SAM heeft meegekregen, maar het benadrukt wel de uitdaging om een goede balans te vinden zodat gebruikers zich niet overvraagd voelen. Daarbij is relevant dat slechts ongeveer 60% van de deelnemers de instructie volgde om een gesprek van 10 minuten te voeren; een aanzienlijk deel ging al na één of twee vragen door naar de vragenlijst, zonder dat er een dialoog over soa’s of condooms had plaatsgevonden. Door de beperkte naleving van de opdracht biedt dit onderzoek helaas ook minder inzicht in de potentiële effectiviteit van SAM in het beïnvloeden van determinanten van condoomgebruik.

Wat betreft de *effectiviteit* toont het onderzoek geen statistisch significante positieve effecten van SAM op de determinanten van condoomgebruik in vergelijking met de controlegroep. Een belangrijke nuancering is dat er sprake was van een significante disbalans in de samenstelling van de groepen: ondanks gerandomiseerde toewijzing bleek de controlegroep bij aanvang al beduidend vaker condooms te gebruiken dan de experimentele groep. Dit verschil kwam pas naar voren bij de data-analyse, omdat condoomgebruik wel in de vragenlijst was opgenomen, maar niet als stratificatiecriterium is gebruikt bij de indeling van de groepen. De verwachting was dat randomisering op leeftijd, geslacht, opleidingsniveau en relatiestatus ook zou zorgen voor een gelijke verdeling van condoomgebruik, maar dat bleek niet het geval. Hierdoor was het te laat om tijdens de dataverzameling nog bij te sturen. Aangezien huidig condoomgebruik een sterke voorspeller is voor de in het experiment gemeten determinanten (zoals ook aangetoond is via aanvullende data-analyses), heeft deze disbalans de resultaten op effectiviteit zeer waarschijnlijk beïnvloed. Daardoor biedt dit experiment beperkt inzicht in de potentiële effectiviteit van SAM, en kunnen geen conclusies worden getrokken over het beïnvloeden van determinanten van condoomgebruik.

Conclusies

Doorontwikkeling van SAM biedt kansen

Uit de co-creatiesessies blijkt dat jongeren een AI-chatbot aantrekkelijk vinden als kanaal voor informatie en advies over seks en relaties. Dit wordt in het onderzoek bevestigd door relatief hoge scores op gebruiksvriendelijkheid en vertrouwen. In de antwoorden op de open vragen in dit

onderzoek, benadrukken jongeren bovendien dat SAM laagdrempelig en anoniem is, waardoor het makkelijker en prettiger wordt om vragen te stellen, ook over gevoelige onderwerpen zoals seksualiteit. Dit vermindert barrières en biedt kansen om content over condoomgebruik en soa-preventie op een natuurlijke en relevante manier te integreren.

Een goed functionerende SAM biedt diverse interessante mogelijkheden. SAM kan jongeren op een directe en laagdrempelige manier voorzien van betrouwbare informatie over seks, relaties en soa-preventie. Daarnaast kan de tool geïntegreerd worden in online campagnes om condoomgebruik te stimuleren. Via social media zoals Instagram, TikTok en Snapchat bereikt SAM jongeren in hun eigen digitale omgeving. Verder kan SAM een waardevolle aanvulling zijn op de seksuele voorlichting vanuit de Centra Seksuele Gezondheid van de GGD.

Betrouwbaarheid als belangrijkste uitgangspunt

Voor gezondheidsinformatie is maximale betrouwbare informatie essentieel. SAM onderscheidt zich hierin van open AI-systemen zoals ChatGPT, die sterk inzetten op dialoogvaardigheden (snelle antwoorden en vloeiend doorvragen) maar minder betrouwbaar zijn omdat er geen volledige transparantie wordt geboden over gebruikte bronnen. SAM kiest bewust voor het tegenovergestelde benadering: maximale betrouwbaarheid (domeingericht en gesloten kennisbank met uitsluitend betrouwbare bronnen), wat ten koste gaat van conversatiekwaliteit.

Usability is goed, maar dialoogkwaliteit moet beter

SAM scoort positief op gebruiksgemak en betrouwbaarheid, maar moet minder robotachtig, sneller en gevarieerder worden om herhalingsgebruik en aanbeveling aan andere jongeren te stimuleren. De AI-assistent van Soatestwijzer.nl wordt over de hele linie iets beter gewaardeerd, vooral door betere conversatievaardigheden. Echter, voor SAM is er bewust voor gekozen om betrouwbaarheid boven gebruiksgemak te stellen. Dankzij een drievoudige controle op interpretatie, antwoordgeneratie en verificatie tegen betrouwbare bronnen is SAM aanzienlijk betrouwbaarder dan de Soatestwijzer.nl AI-assistent en generieke AI-modellen. Dit maakt SAM minder vloeiend in gesprek, maar veel robuuster tegen hallucinaties en misbruik.

Het verschil in usability lijkt echter acceptabel en kan verdwijnen als SAM meer tijd en budget krijgt voor conversational design en uitbreiding van gespreksonderwerpen. Als op basis van dit onderzoek gericht gewerkt wordt aan optimalisering van SAM, kan gestreefd worden naar een livegang op relatief korte termijn. Echte gebruikersdata zijn van groot belang om SAM te trainen, onverwachte fouten te identificeren en continue verbetering van de dialoogkwaliteit te realiseren.

Effectiviteit van SAM nog niet aangetoond

Het onderzoek laat geen significante effecten zien op de geselecteerde gedragsdeterminanten van condoomgebruik. Hiervoor zijn meerdere verklaringen mogelijk. Ten eerste kunnen significante effecten er nog niet zijn. Bijvoorbeeld omdat de beperkte conversatiekwaliteiten van deze eerste versie van SAM de blootstelling aan inhoud over soa's en condooms verminderen. Daarnaast speelt de onderzoeksopzet een rol: bij de randomisatie is niet gestratificeerd naar actueel condoomgebruik. Achteraf blijkt dat deelnemers in de controlegroep significant vaker condooms gebruikten dan deelnemers in de experimentele groep. Omdat actueel condoomgebruik een sterke invloed heeft op gedragsdeterminanten, kan dit verschil de zichtbaarheid van eventuele effecten van SAM substantieel hebben verminderd.

Afsluitende conclusie

De huidige scores weerspiegelen vooral de ontwikkelconditie, niet het potentieel van SAM. De chatbot is ontwikkeld in korte tijd en met beperkt budget, met een complexe taak: niet alleen informatie geven,

maar ook gesprekken voeren over startende relaties, deze richting condoomgebruik sturen en vervolgens doorverwijzen naar betrouwbare bronnen. Dit maakt de vergelijking met andere AI-assistenten, die vooral informatief werken, minder gelijkwaardig.

Daarnaast is gekozen voor een strikte kwaliteitsmethodiek waarbij antwoorden worden gecontroleerd via meerdere modellen en vergeleken met goedgekeurde bronnen. Dit waarborgt betrouwbaarheid, maar beperkt snelheid en variatie. Doorontwikkeling is nodig om dialoogvaardigheden, snelheid en content uit te breiden, zodat SAM zijn rol als schaalbare, laagdrempelige aanvulling op bestaande interventies kan vervullen.

Aanbevelingen

Op basis van de resultaten en conclusies van dit onderzoek worden de volgende aanbevelingen geformuleerd voor de doorontwikkeling van SAM en toekomstig onderzoek.

SAM vormt in potentie een waardevolle aanvulling op de bestaande web-based interventies van Sense: www.sense.info en de Infolijn (telefoon, chat, Whatsapp en email contacten). Waar de website vooral algemene informatie biedt en de Infolijn persoonlijke interactie mogelijk maakt, kan SAM op grote schaal persoonlijk advies geven: duizenden gesprekken tegelijk, 24/7 beschikbaar, geduldig en aanpasbaar aan context en taalniveau. Persoonlijke informatie heeft meer impact dan algemene informatie. SAM kan persoonlijke informatie laagdrempelig (verlaagt barrières om vragen te stellen over gevoelige onderwerpen) aanbieden en vergroot daarmee het bereik, ook via social media. Echter, SAM vervangt de bestaande interventies niet. De Infolijn blijft essentieel voor complexe of emotioneel beladen vragen waarbij menselijke interactie nodig is. Soatestwijzer.nl blijft een betrouwbare bron voor specifieke informatie over soa- en hivtesten. SAM vult deze kanalen aan door jongeren vroeg in hun informatiezoekproces te bereiken en hen door te leiden naar betrouwbare bronnen.

Gegeven het werkdomein seksuele gezondheid, dient maximale betrouwbaarheid van SAM een vereiste te zijn. De aanbeveling is om, met behoud van dit fundament van maximale betrouwbaarheid, de dialoogvaardigheden van de assistent verder te verbeteren. Dit betekent dat de gesprekken vloeiender en menselijker moeten worden gemaakt, en dat de chatbot uitsluitend antwoorden genereert op basis van de geselecteerde, betrouwbare kennisbronnen. Om de effectiviteit van SAM te verbeteren, wordt ook gestreefd naar natuurlijke en vloeiende blootstelling aan soa en condoom content.

Doorontwikkeling van SAM

Korte termijn

De belangrijkste prioriteit voor de komende periode is gericht op het verbeteren van de gebruikerservaring en het investeren in natuurlijke dialogen. Dit kan worden gerealiseerd door meer variatie in de antwoorden aan te brengen en de SAM meer empathisch te laten reageren op de input van gebruikers, mede op basis van de bevindingen uit dit onderzoek. Na de optimalisatieslag kan de effectiviteit van SAM opnieuw worden onderzocht. Bij positieve resultaten van een effectonderzoek is het voornemen SAM op korte termijn te koppelen aan het primaire platform: de website www.sense.info. Echte gebruikersdata is van groot belang om een AI-assistent verder te trainen.

Middellange termijn

Na implementatie op www.sense.info kan de nadruk worden gelegd bij twee kerntaken: borging van de kwaliteit en inhoudelijke opschaling. De kwaliteitsborging van de gebruikerservaring vereist het regelmatig (bijvoorbeeld maandelijks) testen met jongeren en, optioneel, zorgprofessionals, om de

juiste spreektoon te garanderen. Er dient tevens te worden gezorgd voor doorlopend onderhoud van de assistent, waaronder het continu opnieuw trainen van het model en het actualiseren van de kennisbasis.

Daarnaast staat inhoudelijke uitbreiding op de agenda, beginnend met de integratie van multimediale content zoals video's en animaties. Meer blootstelling aan dergelijke voor jongeren aansprekende content kan de effectiviteit van SAM bevorderen. Bij voldoende tijd en budget kan daarnaast ook worden gefocust op het opschalen naar nieuwe onderwerpen (bijvoorbeeld seksuele grensoverschrijding, voorkomen onbedoelde zwangerschappen). Opschaling kan ook gericht zijn op het geschikt maken van SAM voor doelgroepen met een andere seksuele oriëntatie of identiteit.

Toekomstig onderzoek

Voor toekomstig onderzoek verdient het allereerst de aanbeveling om bij randomisatie te stratificeren op actueel condoomgebruik, omdat dit gedrag sterk correleert met de gemeten determinanten. Hierdoor zijn mogelijke effecten van de interventie minder waarneembaar. Een alternatief is om uitsluitend deelnemers te includeren zonder seksuele ervaring, zodat baselineverschillen in condoomgebruik worden vermeden. Daarnaast kan het zinvol zijn om de steekproef te richten op jongeren die laag scoren op de relevante gedragsdeterminanten, omdat deze groep theoretisch meer beïnvloedbaar is en daardoor een groter potentieel effect van de interventie kan laten zien. Verder moet de blootstelling aan de interventie worden vergroot door duidelijke instructies en controlemechanismen, aangezien voor een aanzienlijk deel van de panelrespondenten de vergoeding belangrijker was dan de opdracht volledig uitvoeren.

1 Introductie

1.1 Achtergrond

Het ministerie van Volksgezondheid, Welzijn en Sport (VWS) heeft Soa Aids Nederland in het najaar 2024 de opdracht gegeven om een condoominterventie te ontwikkelen voor de doelgroep jongeren in de leeftijd van 12 – 25 jaar. Aanleiding voor dit verzoek is de stijging van gonorroe infecties in Nederland (RIVM, 2024) en een al jaren zichtbaar dalend condoomgebruik in deze doelgroep (Seks onder je 25^e, 2023). Om deze negatieve trend te keren oriënteerde het Ministerie van VWS zich eveneens op de landelijke inzet op condoompromotie. Omdat onderzoek naar de effectiviteit van een condoominterventie waardevolle inzichten kan opleveren voor de inrichting van een structurele aanpak, is Soa Aids Nederland gevraagd het project in de zomer van 2025 af te ronden. Vanwege deze planning, in combinatie met budgettaire beperkingen in deze fase, is de condoominterventie niet geïmplementeerd en geëvalueerd, maar is gekozen voor het uitvoeren van experimenteel onderzoek naar de acceptatie en effectiviteit van de eerste versie van de interventie. Dit rapport geeft achtereenvolgens inzicht in de achtergronden, doelen en ontwikkeling van de condoominterventie, de methode van (het experimentele) onderzoek en de resultaten. Samenvatting, conclusies en aanbevelingen zijn aan het begin van het rapport opgenomen.

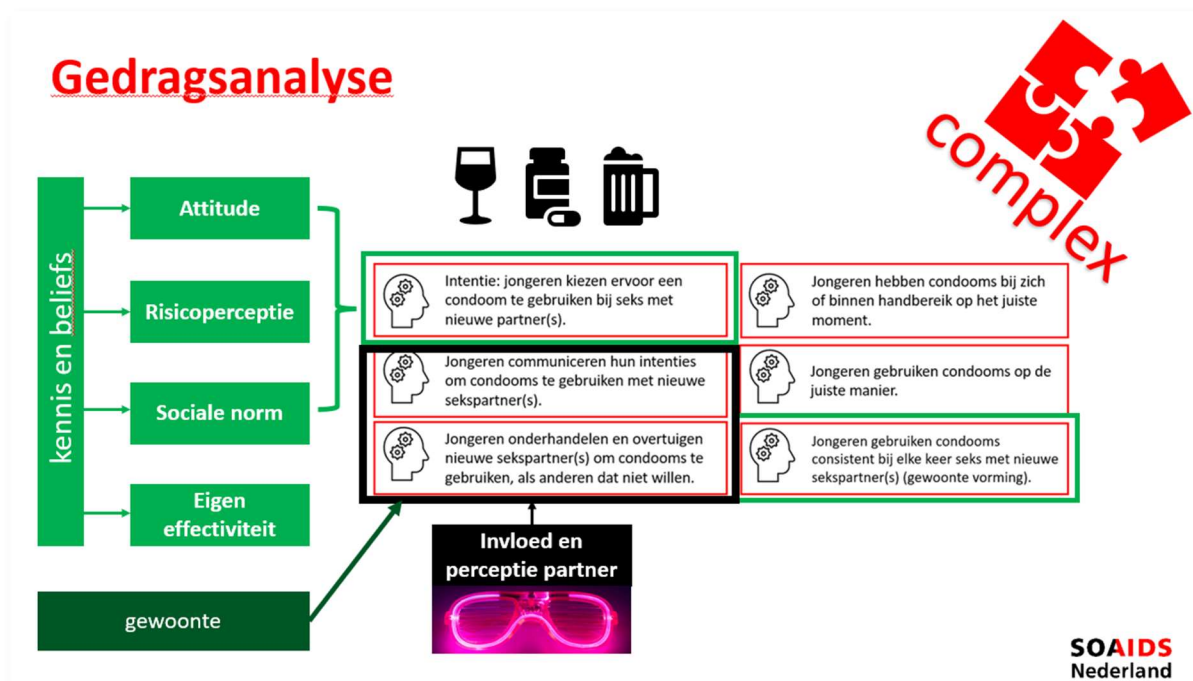
1.2 Aanpak ontwikkeling condoominterventie

Soa Aids Nederland heeft de methode Intervention Mapping toegepast tijdens de ontwikkeling van de condoominterventie. In hoofdlijn bestaat de methode uit het analyseren van het probleem, de doelgroep en de context. Hiervoor is een literatuuronderzoek uitgevoerd naar belangrijke determinanten van condoomgebruik en effectief gebleken condoompromotiestrategieën. Vervolgens zijn veranderdoelen voor de te ontwikkelen condoominterventie opgesteld: welke gedragingen en determinanten moeten veranderen om het gezondheidsdoel te bereiken? Op basis van theoretische onderbouwing worden bewezen effectieve methoden van verandering geselecteerd om ingezet te worden voor interventieontwikkeling. Vervolgens is in co-creatie met jongeren een interventieconcept ontwikkeld dat aansluit bij hun beleavingswereld en communicatievoorkeuren. In het volgende hoofdstuk wordt een samenvatting van de organisatie en resultaten van de co-creatiesessies met jongeren gegeven.

2 Samenvatting resultaten co-creatie met de doelgroep

2.1 Achtergrond uitgangspunten condoominterventie

In figuur 1 zijn de belangrijkste uitkomsten van de gedragsanalyse en belangrijke gedragsdoelen (eerste twee stappen van het Intervention Mapping proces) voor de doelgroep jongeren grafisch weergegeven.

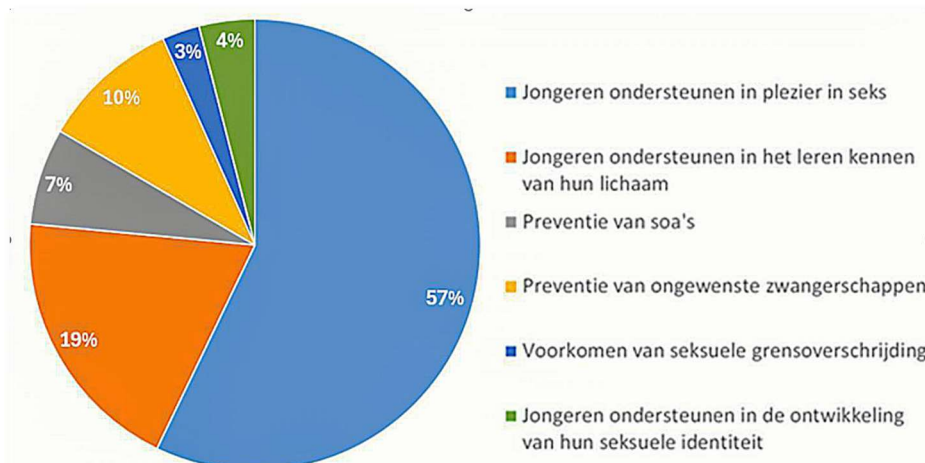


Figuur 1: Samenvatting gedragsanalyse condoomgebruik

De gedragsanalyse¹ laat zien dat condoomgebruik een complex gedrag is, beïnvloed door een breed scala aan factoren. Deze factoren verschillen per individu en situatie en omvatten zowel emotionele, rationele als praktische aspecten. Bovendien is bekend dat positieve intenties om condooms te gebruiken in de praktijk vaak onvoldoende zijn. Situaties zoals het niet beschikbaar hebben van een condoom, middelengebruik (alcohol en drugs) of tegenwerking van een sekspartner kunnen het daadwerkelijke gebruik verhinderen.

Op basis van de gedragsanalyse is besloten een eerste interventie te ontwikkelen die zich richt op het versterken van determinanten die de intentie tot condoomgebruik beïnvloeden: attitude, risicoperceptie, sociale norm en eigen-effectiviteit. Een belangrijke uitdaging voor de acceptatie van deze interventie is dat jongeren in het algemeen weinig spontaan interesse tonen in informatie over condooms en soa's. Dit blijkt ook uit het bezoekgedrag op www.sense.info, de gezamenlijke website over seks en seksuele gezondheid van Soa Aids Nederland, Rutgers en de Centra Seksuele Gezondheid van de GGD-en (circa 7 miljoen paginaweergaven per jaar).

¹ Advies aanpak structurele meerjarige condoompromotie soa preventie, Soa Aids Nederland, Amsterdam (2025)



Figuur 2: Verdeling bezoeken website www.sense.info naar thema

Ruim 75% van de bezoeken op sense.info is gericht op zogenoemde *pleasure-based content* (bijvoorbeeld informatie over seks en relaties), terwijl slechts 25% betrekking heeft op seksuele gezondheidsonderwerpen.

Om de kans op acceptatie van de interventie door jongeren te vergroten, zijn co-creatiesessies georganiseerd (stap 3 van het Intervention Mapping-proces). Het uitgangspunt van deze sessies was het concept van een 'Lovecoach': een digitale coach die jongeren op een toegankelijke en betrouwbare manier advies geeft over onderwerpen waarin zij interesse hebben, zoals daten, relaties en sekstechnieken. Tegelijkertijd stimuleert deze Lovecoach jongeren op een aantrekkelijke manier om aandacht te besteden aan soa-preventie en condoomgebruik. De doelen van de co-creatiesessies waren:

- ✓ Inzicht krijgen in hoe jongeren denken over de uitgangspunten en doelen van de condoominterventie.
- ✓ Mogelijke vervolgstappen bespreken om de interventie zo succesvol mogelijk te ontwikkelen en aan te bieden.
- ✓ Begrijpen welke aanpak, vorm en toon het beste aansluiten bij de doelgroep.

Omdat de eerste versie van de interventie specifiek gericht is op heteroseksuele jongeren, zijn voor de co-creatiesessies uitsluitend jongeren met deze seksuele oriëntatie uitgenodigd.

2.2 Belangrijkste resultaten co-creatiesessies met jongeren

Voor de co-creatiesessies zijn 20 heteroseksuele jongeren gesproken (single- of duo-interviews), met een evenwichtige verdeling naar geslacht (10 jongens, 10 meiden), leeftijd (16–24 jaar) en opleidingsniveau (13 praktisch opgeleid, 7 hoger opgeleid). Alle deelnemers waren niet in een langdurige relatie (>3 maanden).

2.2.1 Belangrijkste bevindingen

Informatiebehoefte:

Jongeren geven aanvankelijk aan weinig behoefte te hebben aan informatie over seksuele gezondheid, maar tijdens de gesprekken blijkt interesse in onderwerpen zoals onzekerheden (wat is normaal?), tips om over seks te praten en betere seks (bijv. standjes, seksspeeltjes, plezier voor beide partners).

Duidelijke voorkeur voor vorm en concept

De jongeren zijn voorbeelden van de Lovecoach in webpagina, app en AI-assistent voorgelegd. Het idee van een AI-chatbot spreekt jongeren unaniem aan. Ze vinden het passend bij deze tijd en nuttig voor zowel algemene als specifieke vragen. De naam Lovecoach roept echter associaties op met relatietherapie en is daarom ongeschikt en moet worden aangepast.

Inhoud en toon:

Jongeren willen korte, duidelijke antwoorden (belangrijk voor praktisch opgeleiden), een uitnodigende toon met suggesties en terugvragen, luchtig maar niet te humoristisch om een gevoel van betrouwbaarheid te behouden. Ervaringsverhalen van andere jongeren en seksweetjes (bijvoorbeeld percentages) zijn aantrekkelijk.

Integratie over seksuele gezondheid:

Informatie over condooms en soa's mag aanwezig zijn, maar subtiel en als onderdeel van bredere thema's, niet te direct.

Gebruiksvorm:

Voorbeeldvragen en categorieën (van luchtig tot serieus) die door de AI-assistent worden aangeboden bij de start en tijdens dialogen, helpen om gesprekken op gang te brengen en te houden.

Bereik:

Sociale media zoals TikTok en Instagram lijken het meest geschikt, maar veel jongeren twifelen ook over het gebruik van de AI-assistent zonder aantrekkelijke en succesvolle promotie. Veel deelnemers kennen sense.info niet.

Samengevat:

Hoewel de jongeren aan het begin van de gesprekken aangaven geen manifeste informatiebehoefte te hebben, wordt het gaandeweg de sessies duidelijk dat ze toch een behoorlijk aantal onderwerpen wel interessant vinden en dat het ook logisch is om seksuele gezondheid hier ook bij te betrekken. Over het algemeen is de doelgroep redelijk positief over het concept, vooral de AI-chatbot, de ervaringsverhalen van anderen en weetjes vinden ze heel aantrekkelijk en interessant. Na de toelichting is het idee van de Lovecoach ook goed, maar de naam moet wel veranderd worden naar iets dat niet direct relatietherapie oproept.

2.2.3 Belangrijkste aanbevelingen uit de co-creatiesessies

- ✓ Een AI-assistent in de vorm van een chatbot verder ontwikkelen als kansrijke interventie, met duidelijke uitleg over mogelijkheden en beperkingen.
- ✓ Maak gebruik van vraagcategorieën en voorbeelden
- ✓ Naam aanpassen naar iets dat uitnodigt en niet aan relatietherapie doet denken.
- ✓ Geen persoonlijkheid of gezicht toevoegen aan de chatbot; dit kan afschrikken.
- ✓ Promotie via sociale media, met prikkelende seksfeitjes, herkenbare ervaringsverhalen en veel voorkomende vragen om jongeren te verleiden tot interactie.
- ✓ Na uitleg over de achtergronden en doelen van sense.info wordt koppeling aan dit platform als goed voor de betrouwbaarheid en als logisch beoordeeld door de jongeren.

2.3 Vervolgstappen

Op basis van de uitkomsten van de co-creatie sessies heeft Soa Aids Nederland – na een aanbestedingstraject waaraan drie gespecialiseerde bureaus hebben deelgenomen - een taakgerichte

AI-Assistent laten ontwikkelen. Deze chatbot, genaamd SAM, heeft als taak om gebruikers te assisteren in het begin van een nieuwe relatie. Daarnaast is SAM geïnstrueerd om gebruikers actief door te geleiden naar betrouwbare online content en interventies die een aantal specifieke en belangrijke determinanten voor condoomgebruik positief beïnvloeden (sociale norm, attitude en risicoperceptie).

Het ontwikkelen van een zogenaamde taakgerichte AI-chatbot voor seksuele gezondheid blijkt ook uit internationaal onderzoek een verstandige keuze omdat het zorgt voor betrouwbare, consistente en doelgerichte ondersteuning bij gevoelige onderwerpen². Open AI-Assistenten zoals ChatGPT hebben een sterke focus op conversationele vaardigheden en kwaliteiten, dit gaat echter te koste van de betrouwbaarheid. SAM neemt een tegenovergestelde benadering; hier ligt de focus op maximale betrouwbaarheid. Tegelijkertijd vraagt dit type chatbot, vergeleken met op open-AI gebaseerde chatbots, om meer tijd en aandacht voor het ontwerpen en bijstellen van dialogen. De interactie moet namelijk niet alleen aansluiten bij de belevingswereld van jongeren, maar ook voldoen aan strenge eisen op het gebied van veiligheid, inclusiviteit en begrijpelijkheid. Gezien de beperkte ontwikkeltijd (twee maanden) richt de eerste versie van SAM zich uitsluitend op heteroseksuele jongeren en op een beperkt aantal gedragsdeterminanten. Bij de selectie en opdrachtverlening is expliciet rekening gehouden met de mogelijkheid om de AI-assistent later op te schalen naar andere doelgroepen en om aanvullende determinanten te integreren.

² Latt PM, Aung ET, Htaik K, Soe NN, Lee D, King AJ, et al. Evaluation of artificial intelligence (AI) chatbots for providing sexual health information: a consensus study using real-world clinical queries. BMC Public Health. 2025 May 15;25(1):1788

3 Methode

3.1 Doel van het experimentele onderzoek

Het experiment heeft twee hoofddoelen:

1. Effectiviteit
Inzicht krijgen in de mate waarin SAM de determinanten van condoomgebruik (attitude, sociale norm, risicoperceptie) positief beïnvloedt.
2. Usability en acceptatie
Inzicht krijgen in hoe gebruiksvriendelijk, betrouwbaar en aantrekkelijk gebruikers de chatbot ervaren.

Om een referentiekader te creëren voor de tweede doelstelling is de op open AI gebaseerde AI-assistent, die wordt ingezet voor ondersteuning van gebruikers van www.soatestwijzer.nl, opgenomen in het experiment. Soatestwijzer.nl is een andere web-based interventie die Soa Aids Nederland ontwikkelt om gebruikers op basis van hun persoonlijke situatie en seksuele gedrag advies te geven over het belang van het doen van een HIV- of soa-test. Door deze opzet van het experiment is het mogelijk de waardering tussen de twee verschillende typen AI-assistenten te vergelijken: taakgericht versus meer open AI-assistent. Tabel 1 geeft een samenvatting van de belangrijkste kenmerken van beide chatbots.

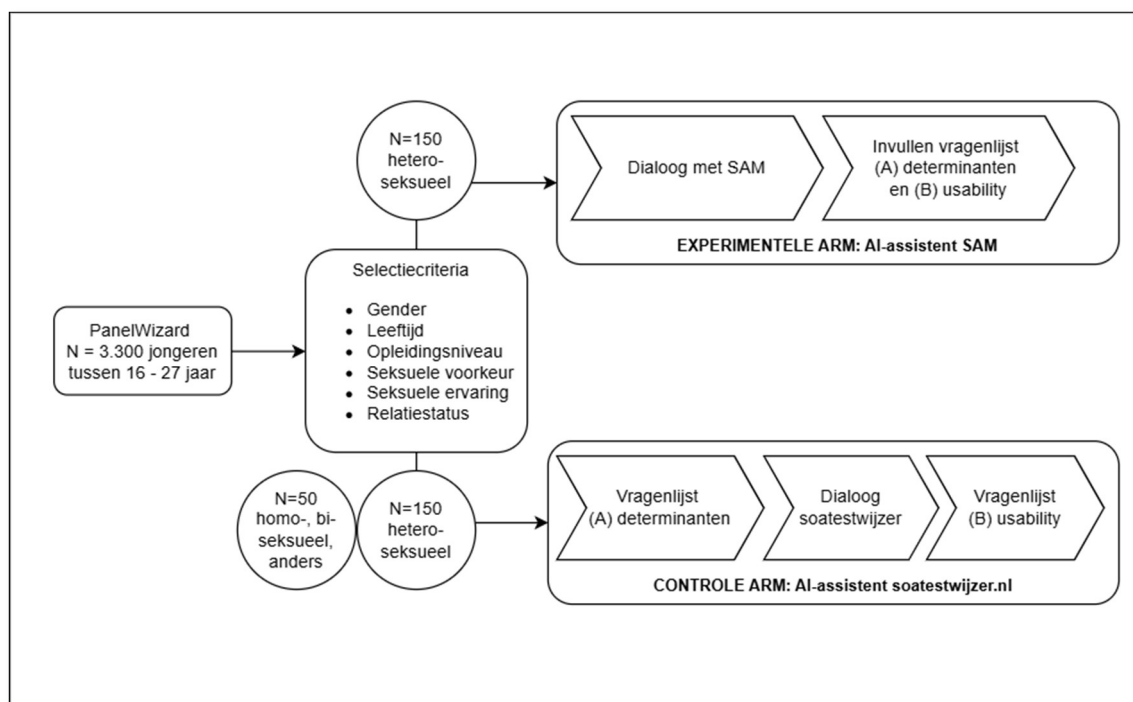
SAM	Chatbot Soatestwijzer.nl
- Taakgericht: duidelijk kader waar de chatbot voor bedoelt is - Opdracht: stuurt dialogen richting een onderwerp, verleidt gebruiker om op een link te klikken - Hoge mate van gebruikersveiligheid en betrouwbaarheid door de drie verschillende LLM's³ (chatbots) - Antwoorden worden altijd vergeleken met content van goedgekeurde bronnen	- De chatbot van Soatestwijzer.nl is gebaseerd op GPT-4 en werkt via een aangepaste interface ('schil') - Via vooraf ingevoerde prompts zijn er regels vastgesteld voor de chatbot: inhoudelijke informatie hoort van soaids.nl te komen. - Meer conversationele ruimte voor de chatbot - In principe is het mogelijk om over meer onderwerpen te praten en 'uit te breken' - Chatbot volgt waar de gebruiker het over wil hebben - Minder betrouwbaarheidsgarantie - Kwaliteit en betrouwbaarheid opschalen is complex en kostbaar

Tabel 1: Samenvatting kenmerken AI-assistenten

3.2 Onderzoeksdesign

In Figuur 3 is de opzet van het experiment schematisch weergegeven.

³ Large Language Models



Figuur 3: Onderzoeksdisegn

Via een werving uit het jongerenpanel van PanelWizard zijn op basis van verschillende selectiecriteria jongeren at random aan twee groepen (experimentele of controle arm) toebedeeld. Om de effectiviteit van SAM te kunnen meten heeft de experimentele arm als eerste opdracht gekregen om een dialoog met SAM te voeren (zie bijlage 1 voor de volledige opdrachtoomschrijving). Na afronding van de dialoog met de AI-assistent zijn de respondenten uit de experimentele arm gevraagd een online vragenlijst in te vullen met zowel het onderdeel A: ‘seksuele gezondheid’ (inclusief vragen over de determinanten van condoomgebruik) als het onderdeel B: ‘usability chatbot’.

Deelnemers aan de controle arm is eerst een vragenlijst met uitsluitend onderdeel A: ‘seksuele gezondheid’ (inclusief vragen over de determinanten van condoomgebruik) voorgelegd. Na afronding van deze vragenlijst is hen gevraagd de chatbot van Soatestwijzer.nl te gebruiken. Na afronding van de dialoog met deze AI-assistent is de vragenlijst met onderdeel B: ‘usability chatbot’ voorgelegd.

Effectiviteit van SAM wordt bepaald door het vergelijken van verschillen op scores van de determinanten van condoomgebruik tussen de experimentele arm (na interactie met SAM) en de controlearm (voor gebruik van Soatestwijzer.nl).

3.3 Respondentselectie

In beide armen is gestreefd naar het realiseren van 150 heteroseksuele mannen en vrouwen omdat de eerste versie van SAM alleen voor deze doelgroep is ontwikkeld. Verder zijn de beide armen zoveel als mogelijk gelijk samengesteld op basis van de volgende (demografische) achtergrondkenmerken:

- Geslacht: man, vrouw
- Opleidingsniveau
- Leeftijd: om voldoende respondenten te kunnen werven is de maximale leeftijd voor deelname aan het experiment opgehoogd naar 28 jaar
- Wel of niet seksuele actief in de afgelopen 6 maanden
- Relatiestatus: geen vaste relatie of een vaste relatie korter dan 2 maanden.

Omdat www.soatestwijzer.nl, anders dan SAM, afgestemd is op andere seksuele oriëntaties dan de heteroseksuele, zijn ook panelleden met verschillende oriëntaties uitgenodigd in de controle arm (N=50). In het vervolg van deze rapportage zijn deze jongeren buiten beschouwing gehouden en zijn alleen de resultaten onder heteroseksuele jongeren uit beide groepen met elkaar vergeleken.

PanelWizard heeft de jongeren uit het panel die aan de selectiecriteria voldoen uitgenodigd voor deelname aan het onderzoek. Zij ontvangen zowel een vergoeding per beantwoorde vraag (€0,10 per vraag) als een vergoeding voor deelname aan het onderzoek (€3,-). Panelleden zijn willekeurig aan de experimentele en controle arm toegewezen, waarbij gelet is dat beide groepen op de verschillende demografische en achtergrondkenmerken gelijk zijn samengesteld.

3.4 Uitvoering onderzoek

Het onderzoek is uitgevoerd aan de hand van de internetmethodiek van PanelWizard. Nadat de doelgroep is geselecteerd op basis van demografische en gedragskenmerken, is een uitnodiging via e-mail of app verstuurd naar de panelleden die aan de criteria voldoen. Respondenten vullen de vragenlijst via hun eigen device zelf online in. Ook het testen van de AI-assistenten is via eigen devices uitgevoerd. Voor SAM hebben de deelnemers aan de experimentele arm een unieke url ontvangen als onderdeel van de onderzoekinstructie. Op basis van de unieke url's kan de inhoud van alle geanonimiseerde dialogen met SAM achteraf worden ingezien en bepaald welke determinanten van condoomgebruik daadwerkelijk aan de orde zijn gekomen. Via het respondentnummer is koppeling met de vragenlijst te maken. Deelnemers aan de controle arm hebben een url van Soatestwijzer.nl ontvangen.

3.5 Inhoud vragenlijst en dataverzameling

De vragenlijst is opgebouwd uit twee onderdelen. In tabel 1 is een overzicht van de inhoud van de vragenlijst gegeven.

Tabel 2: Inhoud vragenlijst

Onderdeel Seksuele Gezondheid (Determinanten)	Onderdeel Usability van de chatbot
- Hoe oud was u toen u voor het eerst vaginale seks (penis in vagina) had? - Hoe vaak gebruikt u een condoom bij vaginale seks (penis in vagina)? - Hoeveel vertrouwen heeft u erin...? - dat u weet hoe u een condoom gebruikt? (bijv. uit verpakking halen, omdoen) - dat u een (nieuwe) (seks)partner kan overtuigen om een condoom te gebruiken? - Stel dat u seks gaat hebben met een nieuwe (seks)partner. In hoeverre bent u van plan om een condoom te gebruiken in onderstaande situaties? Ik ben van plan om een condoom te gebruiken, als ik seks heb met een nieuwe (seks)partner - ook als ik en mijn nieuwe (seks)partner elkaar vertrouwen - ook als ik en mijn nieuwe (seks)partner al een andere manier gebruiken om niet zwanger te worden	- In hoeverre bent u het eens met onderstaande uitspraken over de chatbot¹? (helemaal oneens-helemaal eens) - De persoonlijkheid van de chatbot was realistisch en boeiend - De chatbot leek te robotachtig - De chatbot was verwelkomend bij het eerste gebruik - De chatbot kwam erg onvriendelijk over - De chatbot begreep mij goed - De chatbot herkende veel van mijn invoer niet - In hoeverre bent u het eens met onderstaande uitspraken over de <i>gebruiksvriendelijkheid</i> van de chatbot¹? (helemaal oneens-helemaal eens) - De chatbot legde zijn doel en functie goed uit - De chatbot gaf geen enkele aanwijzing over zijn doel - De chatbot was gemakkelijk te verkennen - De chatbot was erg makkelijk om te gebruiken - De chatbot was erg complex

5. Hieronder ziet u enkele tegenstellingen over het gebruik van condooms met een nieuwe (seks)partner. Wat vindt u van het gebruik van condooms met een nieuwe (seks)partner?⁴ a) Heel onplezierig- heel plezierig b) Helemaal niet fijn-heel fijn c) Helemaal geen gedoe-veel gedoe d) Heeft veel nadelen-heeft veel voordelen e) Heel onverstandig-heel verstandig 6. Hoeveel mensen in uw omgeving denkt u dat een condoom gebruiken¹? a) Bijna niemand – (vrijwel) iedereen 7. Hoeveel mensen in uw omgeving denkt u dat een condoom gebruiken als ze seks zouden hebben met een nieuwe (seks)partner¹? a) Bijna niemand – (vrijwel) iedereen 8. Weet u wat een soa is? (ja/nee) 9. Hoe erg zou u het vinden om een soa te krijgen¹? a) helemaal niet erg-heel erg 10. Heeft u ooit een soa gehad? Zo ja, welke? 11. Hoe groot schat u de kans in dat u in de komende 6 maanden een soa oploopt¹? a) Zeer klein-heel groot 12. Denkt u dat onderstaande uitspraken waar of niet waar zijn? (waar, niet waar, weet niet) a) Een soa kun je oplopen als je één keer seks hebt zonder condoom b) Als je je na seks wast, krijg je minder snel een soa c) Als je geen lichamelijke klachten hebt, kun je toch een soa hebben d) Je kunt een soa krijgen van iemand die zelf geen klachten heeft. 13. In hoeverre zijn onderstaande aspecten belangrijk voor u¹? (heel onbelangrijk/heel belangrijk) a) Seksuele gezondheid b) Tijdens seks met een nieuwe (seks)partner een condoom gebruiken c) Ervoor zorgen dat ik geen soa krijg 14. Gelden onderstaande uitspraken wel of niet voor u? (Wel, niet, weet niet) a) Mensen die belangrijk zijn voor mij (zoals vrienden en familie) vinden dat ik een condoom moet gebruiken als ik seks heb met een nieuwe (seks)partner b) Mensen die belangrijk zijn voor mij (zoals vrienden of klasgenoten) gebruiken zelf condooms als ze seks hebben met een nieuwe (seks)partner c) Als ik een condoom gebruik, vertrouw ik erop dat ik beschermd ben tegen soa's	f) Het was makkelijk om in de war te raken bij het gebruik van de chatbot 3. In hoeverre bent u het eens met onderstaande uitspraken over <i>de antwoorden</i> van de chatbot¹? (helemaal oneens-helemaal eens) a) De reactiesnelheid van de chatbot was goed b) De antwoorden van de chatbot waren nuttig, passend en informatief c) De antwoorden van de chatbot waren irrelevant d) De chatbot ging goed om met vergissingen of fouten e) De chatbot kon niet omgaan met fouten 4. In hoeverre bent u het eens met onderstaande uitspraken over de chatbot¹? (helemaal niet mee eens-helemaal mee eens) a) De chatbot begon veel vaker over condooms en/of seksuele gezondheid dan ik wilde b) Als ik weer een vraag heb over seksualiteit of gezondheid, zou ik deze aan de chatbot stellen c) Ik vertrouw erop dat de chatbot informatie van betrouwbare bronnen gebruikt d) Ik zou deze chatbot aanraden aan mensen in mijn omgeving als zij een vraag hebben over seksuele gezondheid 5. Wat vindt u goed aan de chatbot? (open) 6. Wat zou volgens u verbeterd moeten worden aan de chatbot? (open) 7. In hoeverre bent u het eens met onderstaande uitspraken over technologie in het algemeen¹? a) Ik geloof in het algemeen dat technologie dingen beter maakt b) Ik ben meestal een van de eersten die nieuwe technologieën uitprobeert c) Ik voel me ongemakkelijk als ik nieuwe technologieën moet leren d) Ik maak me zorgen dat het gebruik van nieuwe technologieën tot fouten kan leiden
--	--

In de periode 28 augustus 2025 tot en met 19 september 2025 heeft de dataverzameling plaatsgevonden.

⁴ Gemeten op een 5-puntschaal

4 Resultaten

4.1 Respons

Tabel 2 toont de gerealiseerde respons van zowel de experimentele als de controle arm met de achtergrondkenmerken.

Tabel 3: Gerealiseerde respons en achtergrondkenmerken

	Experimentele arm	Controle arm
Bruto steekproef	268	250
Verwijderd na datacheck	1	1
Afgebroken vragenlijsten	40	8
Geen medewerking	25	37
Afgeronde vragenlijsten	202	204
Geslacht (%)		
Man	33,9	33
Vrouw	66,1	67
Leeftijd (%)		
16 t/m 19 jaar	22,8	25,4
20 t/m 25 jaar	46,2	43,7
26 t/m 29 jaar	31,0	30,9
Opleidingsniveau (%)		
○ (Geen onderwijs) / basisonderwijs / lagere school	4,5	5,7
○ LBO / VBO / VMBO	5,0	4,2
○ MAVO / eerste 3 jaar HAVO en VWO / VMBO	7,9	9,4
○ MBO	19,4	15,1
○ HAVO en VWO bovenbouw / WO-propedeuse	27,7	30,2
○ HBO / WO-bachelor of kandidaats	24,0	27,8
○ WO-doctoraal of master	11,6	7,5
Seksuele oriëntatie	Heteroseksueel	Divers
Seksueel actief afgelopen 6 maanden		
Ja	40,5	38,2
Nee	59,5	61,8

Respondenten zijn wat betreft de verschillende achtergrondkenmerken redelijk gelijk verdeeld over de twee onderzoeksgroepen.

In paragraaf 4.2 worden de belangrijkste uitkomsten van het onderzoek gepresenteerd die betrekking hebben op de beoordeling van de effectiviteit van SAM. In dit deel van het rapport zijn alleen de data gebruikt van de respondenten in beide onderzoeksgroepen die vallen binnen de doelgroep van SAM: heteroseksueel en jonger dan 25 jaar. Bij de bespreking van de resultaten van de usability van de AI-assistenten in paragraaf 4.3 zijn alle respondenten aan het onderzoek weer geïnccludeerd.

4.2 Beoordeling effectiviteit AI-assistent SAM

4.2.1 Inleiding

SAM is ontwikkeld met als doelstelling de determinanten van condoomgebruik te beïnvloeden. Uit de gedragsanalyse komt naar voren dat de determinant 'intentie van condoomgebruik' een belangrijke voorspeller van condoomgebruik is. De intentie zelf wordt ook beïnvloed door verschillende determinanten. De eerste versie van SAM is geprogrammeerd om de determinanten sociale norm, attitude ten aanzien van condoomgebruik, eigen-effectiviteit en de risicoperceptie (ernst en kans van het oplopen van een soa) te bespreken en positief te beïnvloeden.

De effectiviteit van SAM is geëvalueerd door de scores op geselecteerde gedragsdeterminanten te vergelijken tussen de experimentele en controlegroep. Tenzij anders vermeld, zijn deze determinanten gemeten met behulp van een 5-puntsschaal. In deze paragraaf worden eerst de gemiddelde scores van alle respondenten binnen de doelgroep (heteroseksueel en jonger dan 25 jaar) op de verschillende determinanten gepresenteerd. Indien er verschillen tussen de experimentele en controlegroep worden waargenomen, zijn aanvullende statistische analyses uitgevoerd om te bepalen of deze verschillen significant zijn.

4.2.2 Kwaliteit van de dialogen en respondentselectie

De evaluatie omvatte tevens een analyse van de kwaliteit van de gevoerde dialogen met de geautomatiseerde assistent (SAM). Conform de uitgangspunten uiteengezet in Hoofdstuk 3, was SAM geprogrammeerd om gespreksthema's die spontaan door de doelgroep geïnitieerd werden (zoals relaties en sekstechnieken) vloeiend te overbruggen naar onderwerpen gerelateerd aan soa's en determinanten van condoomgebruik.

Het realiseren van deze vloeiende gespreksverschuiving blijkt een complexe taak en is niet in alle dialogen succesvol gebleken.

Op basis van een handmatige inventarisatie van de dialogen is vastgesteld dat:

- Van de totale groep van N=120 respondenten in de experimentele conditie (die voldoen aan de gespecificeerde leeftijdscategorie), heeft een subgroep van n=71 respondenten daadwerkelijk aandacht besteed aan de onderwerpen soa's en condoomgebruik in hun dialoog met SAM.
- Dit komt neer op 59.2% (71/120) van de dialogen in de experimentele arm.

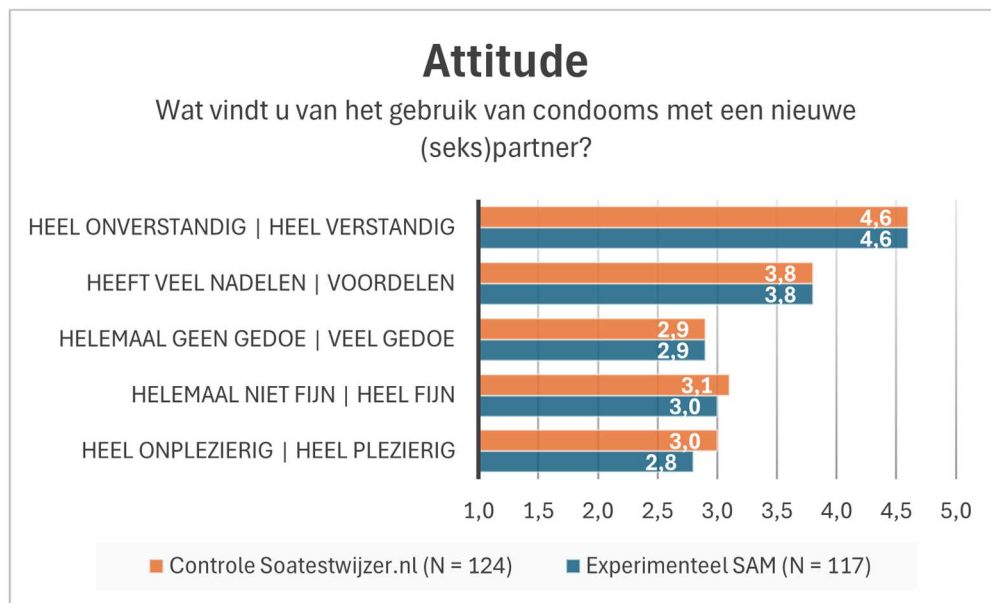
Aangezien de verwachte effectiviteit van de interventie afhankelijk is van de succesvolle blootstelling aan de beoogde inhoud (over soa's en condooms), kunnen significante effecten van SAM uitsluitend bij deze geselecteerde groep (N=71) worden verwacht. Afgezien van de leeftijd – deze subgroep is gemiddeld iets jonger dan de oorspronkelijke onderzoeksgroepen – zijn er verschillen waarneembaar op achtergrondkenmerken. Daarom zijn er voor deze subgroep aanvullende analyses uitgevoerd om te onderzoeken of op de gemiddelde scores van de determinanten significant verschillen worden waargenomen met de controle groep.

Voor de vier specifieke determinanten (attitude, sociale norm, eigen-effectiviteit en risicoperceptie) is een aanvullende handmatige analyse van de dialogen met SAM uitgevoerd om te beoordelen in hoeverre binnen de experimentele groep expliciet aandacht aan deze determinanten is besteed. Gezien de beperkte omvang van deze subgroepen (attitude (n = 18), sociale norm (n = 4), eigen-

effectiviteit (n= 12) en risicoperceptie (n=14) en de daarmee samenhangende lage statistische betrouwbaarheid, zijn er voor deze subgroepen geen aanvullende statistische analyses uitgevoerd.

4.2.3 Scores op determinanten

De determinant attitude is zowel op affectief (gedoe, fijn, plezierig) en instrumenteel niveau (verstandig, voordelen) gemeten, zie grafiek 1.

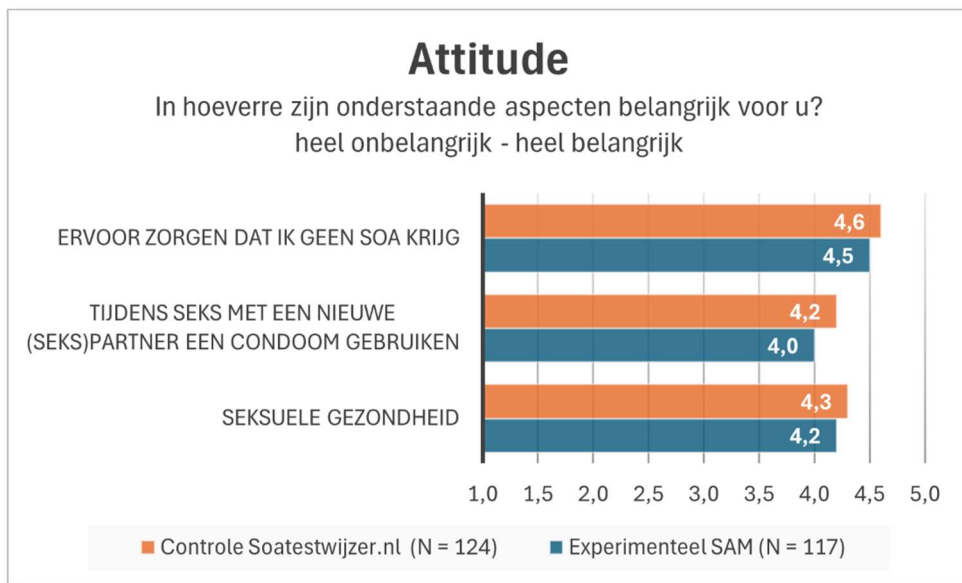


Grafiek 1: Affectieve en instrumentele attitude t.a.v. condoomgebruik

Er zijn slechts beperkte verschillen zichtbaar op de gemiddelde scores van de affectieve en instrumentele attitude tussen de twee onderzoeksgroepen. Op de affectieve attitude (fijn, plezier) scoort de controle arm (fijn: $M_{\text{controle}}=3,1$, $M_{\text{experiment}}=3,0$ | plezierig: $M_{\text{controle}}=3,0$, $M_{\text{experiment}}=2,8$) net iets beter dan de experimentele arm. Uit de aanvullende analyse blijkt dat de verschillen op de affectieve attitude 'plezier' significant is: $t(276) = -2.15$, $p = .032$. Echter, als we inzoomen op de subgroep die in de dialoog met SAM in ieder geval aandacht aan soa's en condooms heeft besteed en deze vragen hebben beantwoord ($N=49$)⁵, dan zijn er geen significante verschillen meer tussen twee onderzoeksgroepen. Onafhankelijke t-toetsen om verschillen tussen de twee groepen op alle affectieve en instrumentele attitude-items tonen geen significant verschillen aan ($p > .05$).

Naast attitudes ten aanzien van condoomgebruik zijn ook de attitudes ten aanzien van seksuele gezondheid en condoomgebruik in een nieuwe relatie onderzocht. In grafiek 2 staan de resultaten samengevat.

⁵ Soms is deze subgroep kleiner dan $N=71$ omdat er geen antwoord door de respondent is ingevuld of gebruik is gemaakt van de categorie 'zeg ik liever niet'.

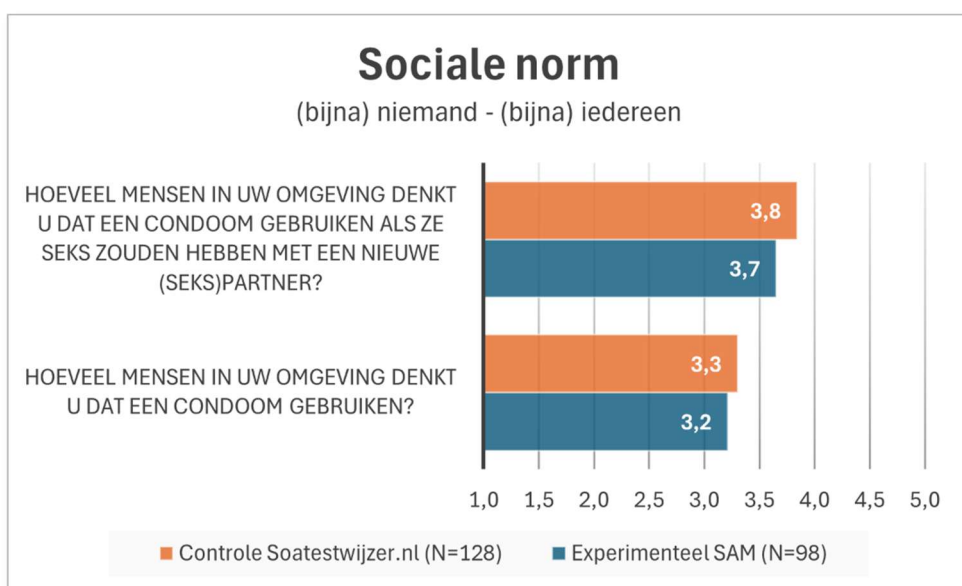


Grafiek 2: Attituden t.a.v. seksuele gezondheid en condoomgebruik in een nieuwe relatie

Op alle drie de items zijn de gemiddelde scores van de respondenten uit de controle arm iets hoger in vergelijking met de experimentele groep. Over het algemeen is seksuele gezondheid en het voorkomen van soa's voor beide groepen belangrijk met gemiddelde scores boven vier op een 5-puntschaal. De controle groep scoort op alle items iets positiever, alleen voor het item 'condoomgebruik met een nieuwe (seks)partner', is het verschil op de gemiddelde scores: ($M_{\text{controle}}=4.2$, $M_{\text{experiment}}=4.0$) significant: $t(269) = -1.97$, $p = .049$.

Onafhankelijke t-toetsen voor vergelijking tussen de subgroep uit de experimentele arm die daadwerkelijk een dialoog met SAM over soa's en condooms heeft gevoerd en deze vragen hebben beantwoord ($N=47$) en de controle arm, tonen aan dat er geen significante verschillen op de drie attitude-items waarneembaar meer zijn ($p > .05$).

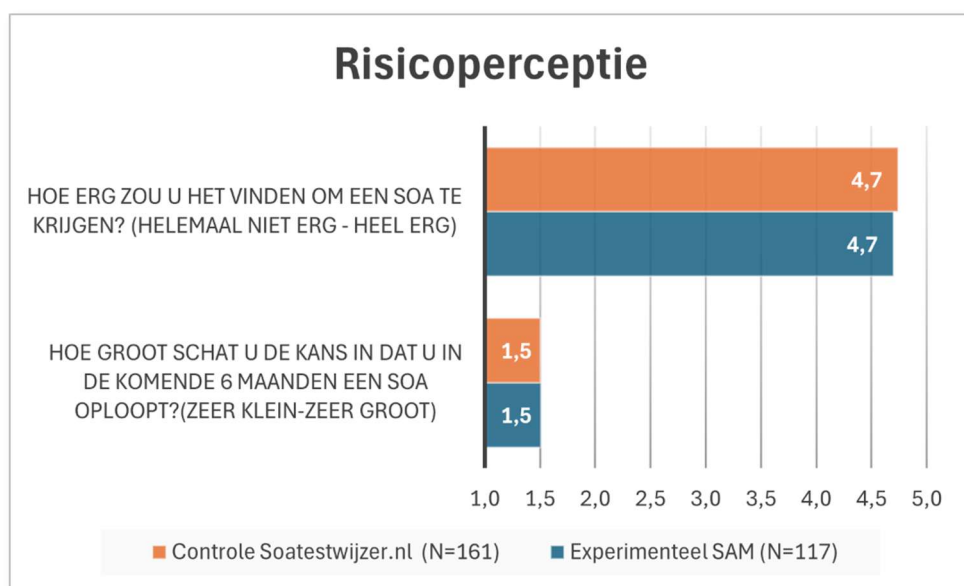
In grafiek 3 zijn de scores opgenomen op de items die de sociale norm meten.



Grafiek 3: Sociale norm

De gemiddelde scores op de items die de sociale norm meten zijn in beide groepen ongeveer gelijk. De controle arm scoort voor beide items net iets hoger. Aanvullende analyses tonen aan dat de verschillen op de gemiddelde scores niet significant zijn ($p > .05$). Ook als we de gemiddelde scores van de subgroep die daadwerkelijk een dialoog met SAM heeft gevoerd over soa's en condoomgebruik en deze vragen hebben beantwoord (N=39 voor item 'condoomgebruik', N=49 voor item 'met een nieuwe sekspartner') vergelijken met de controle groep, zijn er geen significante verschillen waarneembaar ($p > .05$).

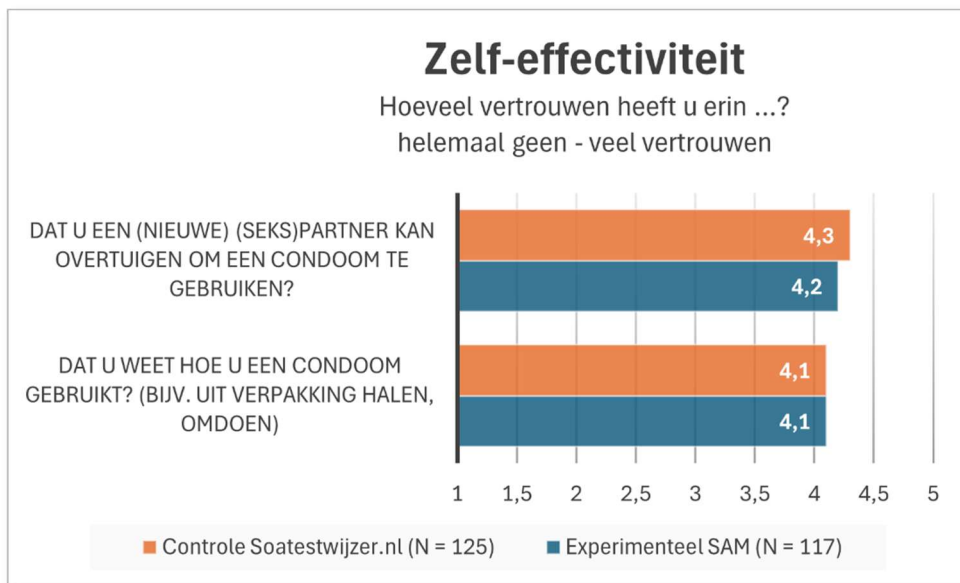
In grafiek 4 zijn de gemiddelde scores voor twee dimensies van de risicoperceptie weergegeven. Het gaat om een inschattingen betreffende de ernst en de kans een soa op te lopen.



Grafiek 4: Risicoperceptie

De gemiddelde scores op beide items zijn gelijk voor beide onderzoeksgroepen. Respondenten in beide groepen vinden het erg om een soa te krijgen maar schatten tegelijkertijd de kans dat zij een in het komende half jaar een soa oplopen als gering in. Aanvullende analyses wijzen uit dat er geen significante verschillen tussen beide groepen waarneembaar zijn ($p < .05$). Ook als we de gemiddelde scores van de subgroep die daadwerkelijk een dialoog met SAM heeft gevoerd over soa's en condoomgebruik vergelijken met de controle groep (N=49), zijn er geen significante verschillen waarneembaar ($p > .05$).

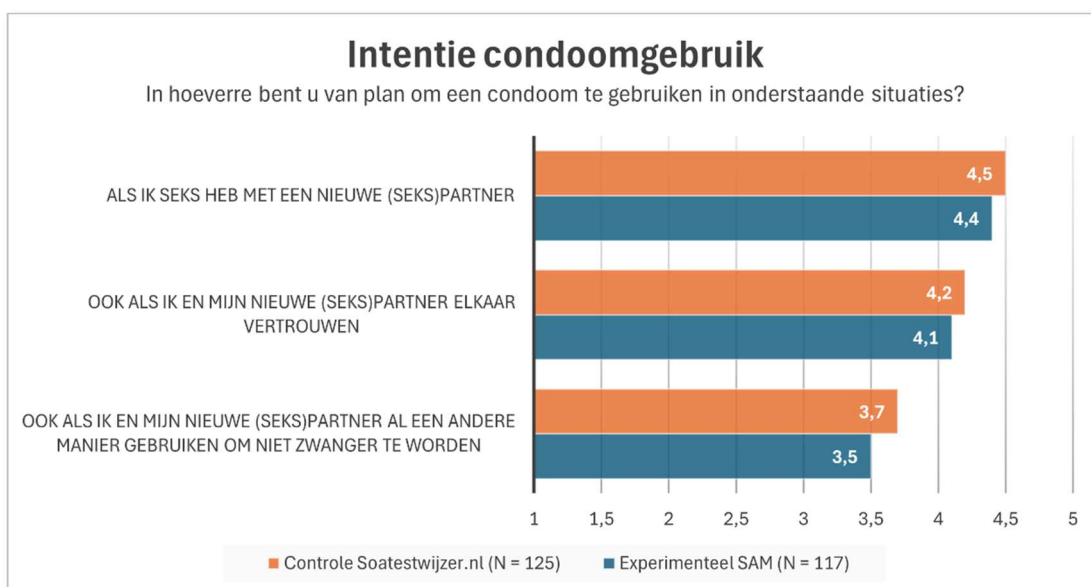
De zelf-effectiviteit (mate waarin een persoon eigen vaardigheden inschat) is zowel gemeten voor communicatie en onderhandelingsvaardigheden als praktische vaardigheden om een condoom correct te kunnen gebruiken. De gemiddelde scores op deze items zijn opgenomen in grafiek 5.



Grafiek 5: Zelf-effectiviteit condoomgebruik

In beide armen zijn de respondenten positief over hun eigen vaardigheden. De gemiddelde scores op beide items wijzen op een redelijk hoog zelfvertrouwen. Aanvullende analyses wijzen uit dat er geen significante verschillen waarneembaar zijn tussen de gemiddelde scores in beide onderzoeksgroepen ($p < .05$). Ook als we de gemiddelde scores van de subgroep die daadwerkelijk een dialoog met SAM heeft gevoerd over soa's en condoomgebruik vergelijken met de controle groep ($N=49$), zijn er geen significante verschillen waarneembaar ($p > .05$).

Volgens de uitgevoerde gedragsanalyse beïnvloeden de onderzochte determinanten de intentie van condoomgebruik. De intentie zelf is in het onderzoek voor drie verschillende situaties gemeten en zijn in grafiek 6 opgenomen.

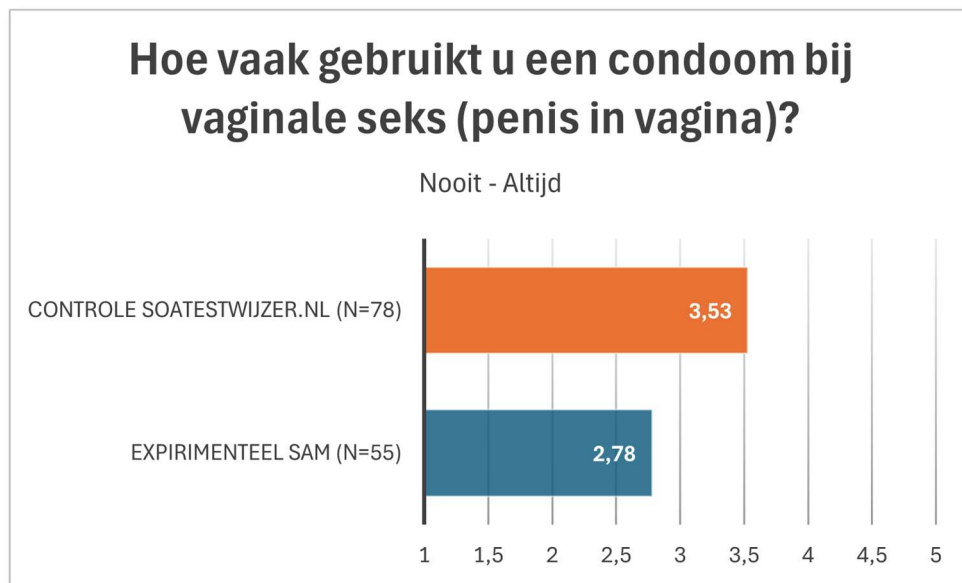


Grafiek 6: Intentie condoomgebruik

De gemiddelde scores op de intentie om een condoom te gebruiken met een nieuwe sekspartner is relatief hoog in beide groepen voor de verschillende situaties. Vertrouwen in een nieuwe sekspartner

heeft enige invloed en verlaagt de intentie om een condoom te gebruiken enigszins. De invloed van het 'gebruik van al een andere manier van anticonceptie' op de intentie een condoom te gebruiken is groter en vertaalt zich in de laagste gemiddelde scores op de intentie voor beide onderzoeksgroepen. Aanvullende analyses tonen geen significante verschillen aan tussen de gemiddelde scores op de drie gemeten situaties voor beide groepen ($p < .05$). Ook de gemiddelde scores van de subgroep die daadwerkelijk een dialoog met SAM heeft gevoerd ($N=49$) over soa's en condoomgebruik wijken niet significant af van de gemiddelde scores van de controle groep ($p > .05$).

Daarnaast is het actuele condoomgebruik van de respondenten in kaart gebracht. Grafiek 7 toont de gemiddelde scores per onderzoeksgroep. Bij deze vraag was de antwoordoptie 'zeg ik liever niet' beschikbaar, waardoor de omvang van de subgroepen relatief beperkt is.



Grafiek 7: Condoomgebruik

Er is een onafhankelijke t-toets uitgevoerd om te bepalen of de verschillen voor de gemiddelde scores op condoomgebruik tussen de experimentele en controlegroep significant zijn. De resultaten tonen een significant verschil tussen de groepen, $t(131) = -2.98$, $p = .003$. Dit betekent dat de experimentele groep een lagere gemiddelde score had op condoomgebruik dan de controlegroep.

Deze bevinding is relevant voor de interpretatie van de onderzoeksresultaten. Uit de gedragsanalyse blijkt dat gewoontegedrag een belangrijke voorspeller is van condoomgebruik. Hoewel de experimentele en controlegroep vergelijkbaar zijn samengesteld op demografische kenmerken (leeftijd, opleidingsniveau, geslacht en seksuele voorkeur), is er niet gestratificeerd op actueel condoomgebruik. Hierdoor is het aannemelijk dat actueel condoomgebruik invloed heeft op de scores van de onderzochte determinanten. Dit kan ertoe leiden dat het effect van de interventie in de experimentele groep minder duidelijk zichtbaar is.

Om de individuele invloed van condoomgebruik op de vier gemeten gedragsdeterminanten (attitude, subjectieve norm, waargenomen eigen-effectiviteit en risicoperceptie) te onderzoeken, en daarmee de potentiële bijdrage van deze determinanten aan eventuele groepsverschillen (experimentele versus controlegroep) in relatie tot de invloed van SAM, zijn univariate analyses van variantie (ANOVA's) uitgevoerd.

Hoewel de analyses geen statistisch significante verschillen aantoonde tussen de experimentele en de controlegroep op elk van deze determinanten ($P > .05$), blijkt uit de resultaten dat condoomgebruik een sterke en consistente voorspeller is van alle gemeten determinanten.

4.2.4 Conclusies effectiviteit condoominterventie SAM

Op basis van de in de vorige paragraaf 5.2.3 gepresenteerde resultaten van het onderzoek is de effectiviteit van SAM niet aangetoond. Er zijn maar kleine verschillen waarneembaar tussen de experimentele en controlegroep op de gemiddelde scores van de determinanten. Als zich verschillen voordoen, dan scoort de controlegroep iets hoger dan de experimentele groep die is blootgesteld aan SAM. Aanvullende analyses hebben aangetoond dat deze kleine verschillen niet significant zijn.

Het ontbreken van significante verschillen tussen de experimentele en controlegroep op de determinanten impliceert dat de SAM-interventie er niet in is geslaagd om deze determinanten effectief te beïnvloeden. Mogelijke verklaringen hiervoor zijn:

- ✓ Blootstelling:
De interventie was mogelijk niet langdurig genoeg om zichtbare verschillen op determinanten te realiseren. Op basis van de handmatige inventarisatie blijkt slechts ongeveer 60% van de experimentele groep daadwerkelijk een gesprek over soa's en/of condooms met SAM te hebben gevoerd. Een deel van de respondenten heeft niet aan de opdracht voldaan en slechts één of twee vragen aan SAM gesteld en is vervolgens de vragenlijst gaan invullen.
- ✓ Inhoud:
Bij respondenten die een geslaagde dialoog met SAM hebben gevoerd (in termen van voldoende blootstelling), hoeft de inhoud van die dialoog niet altijd aan te sluiten bij hun specifieke behoefte. Zo kan de dialoog bijvoorbeeld gericht zijn geweest op soa's en instrumentele attitude, terwijl voor een bepaalde respondent juist de affectieve attitude belangrijker was en niet aan bod kwam.
- ✓ Veranderbaarheid van de determinant:
De determinanten (bijvoorbeeld instrumentele attitude en risicoperceptie) waren in de uitgangssituatie al op redelijk hoog niveau, waardoor er beperkte ruimte was voor verdere verbetering door de interventie.

Echter, vanwege de opzet van het experimentele onderzoek is het niet uit te sluiten dat de invloed van SAM wordt vertekend door verschillen in actueel condoomgebruik tussen de onderzoeksgroepen. De gemiddelde score op condoomgebruik ligt significant hoger in de controlegroep dan in de experimentele groep. Univariate analyses tonen aan dat verschillen in condoomgebruik een substantieel deel van de variantie in de vier gemeten determinanten verklaren. Dit sterke verband komt overeen met theoretische modellen, zoals de Theory of Planned Behavior, en bevestigt dat recent gedrag (condoomgebruik) sterk correleert met determinanten van condoomgebruik, zoals ook besproken in de uitgevoerde gedragsanalyse.

In toekomstig onderzoek is het aan te bevelen om bij randomisatie te stratificeren op actueel condoomgebruik, zodat beide onderzoeksgroepen vergelijkbaar zijn. Een alternatief is om uitsluitend respondenten te includeren die nog geen seksuele (penis in vagina) ervaring hebben opgedaan, om baselineverschillen in condoomgebruik te vermijden. Verder kan overwogen worden om uitsluitend respondenten te includeren die laag scoren op de relevante gedragsdeterminanten (zoals attitude, sociale norm, eigen-effectiviteit en risicoperceptie). Deze groep is waarschijnlijk meer beïnvloedbaar, waardoor de interventie een groter potentieel effect kan hebben.

4.3 Beoordeling usability chatbots

4.3.1 Inleiding

In deze paragraaf worden de resultaten van de vragenlijst over de 'usability van de chatbots besproken: het gaat om SAM binnen de experimentele groep en Soatestwijzer.nl binnen de controlegroep. Omdat de beoordeling op veel aspecten niet direct afhankelijk is van de inhoud van de dialoog, worden eerst de resultaten van alle respondenten uit beide onderzoeksgroepen gepresenteerd. Daarnaast zijn ook opgenomen de gemiddelde scores van de respondenten uit de experimentele groep die op basis van de handmatige inventarisatie een geslaagde dialoog met SAM hebben gevoerd (N=93). Respondenten met minder dan 5 vragen en antwoorden zijn uitgesloten. In de grafieken is de groep aangeduid als 'geslaagde dialoog'. Hier zitten dus ook dialogen (N=22) bij die alleen over relaties en seksualiteit zijn gegaan en niet over soa's en condooms.

Omdat het in dit onderdeel van het onderzoek gaat om gebruikservaring en niet om het effect van de interventie of hypothese toetsing, zijn niet voor alle gemiddelde scores de resultaten van aanvullende analyses gepresenteerd om eventuele significante verschillen tussen de groepen aan te tonen. De resultaten zijn primair bedoeld om inzicht te geven in de gebruikservaring en verbeterpunten. Daarom ligt de nadruk op het beschrijven van gemiddelde scores en opvallende patronen.

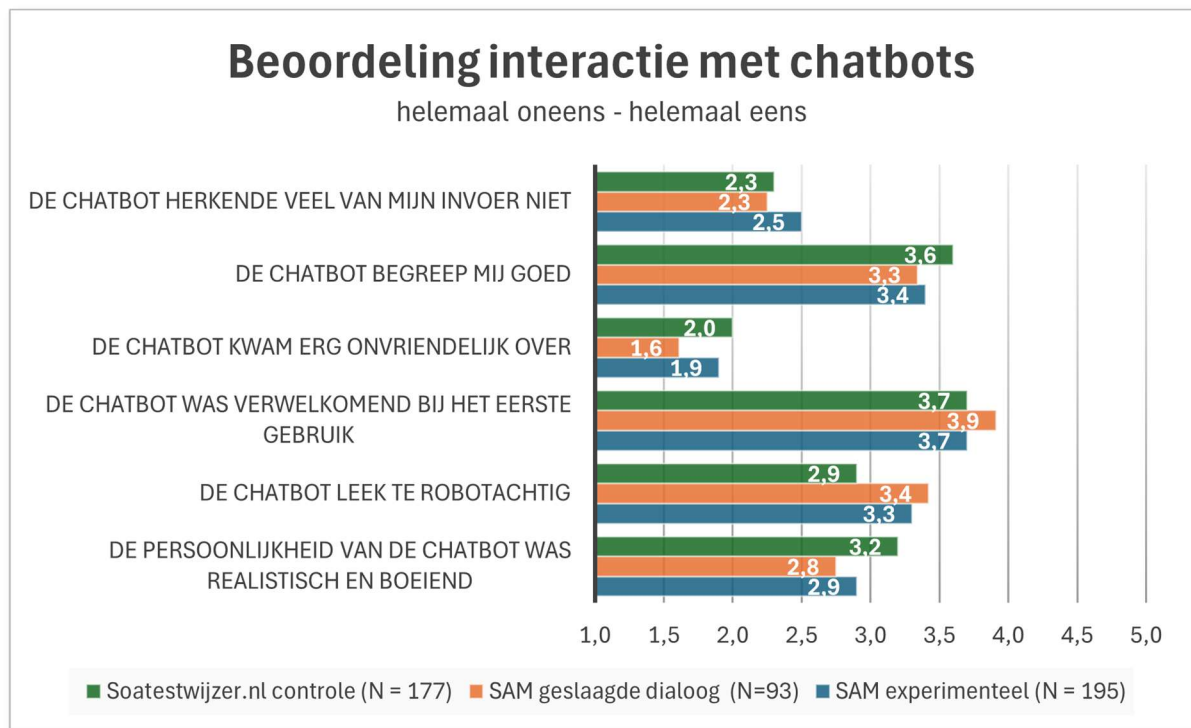
Om te onderzoeken of de houding ten opzichte van nieuwe technologie van invloed is op de gemiddelde scores van de onderzoeksgroepen, zijn vier stellingen opgenomen in de vragenlijst. Respondenten beoordeelden deze op een 5-puntschaal (helemaal niet mee eens – helemaal mee eens). De stellingen meten algemene technologische acceptatie, innovativiteit en mogelijke terughoudendheid:

- a) Ik geloof dat technologie in het algemeen dingen beter maakt
- b) Ik probeer meestal als een van de eersten nieuwe technologieën uit
- c) Ik voel me ongemakkelijk als ik nieuwe technologieën moet leren
- d) Ik maak me zorgen dat het gebruik van nieuwe technologieën tot fouten kan leiden

Voor stellingen c en d zijn de waarden omgekeerd gecodeerd, zodat een hogere score een positievere houding tegenover technologie weergeeft. Daarna is een samengestelde variabele berekend met de gemiddelde score per respondent. De experimentele groep scoorde ($M = 3,33$) en de controlegroep ($M = 3,24$), wat wijst op een vergelijkbare, licht positieve houding in beide groepen. Een onafhankelijke t-test liet geen significant verschil zien: $t(271) = 1,01$, $p = 0,314$. Dit suggereert dat technologische acceptatie niet verschilt tussen de groepen en waarschijnlijk geen verklaring biedt voor verschillen in chatbotbeoordeling.

4.3.2 Beoordeling interactie

In grafiek 8 worden de resultaten van de beoordeling van verschillende aspecten van de interactie met de AI-assistenten gepresenteerd. De interactie is met vijf stellingen gemeten ieder op een 5-puntsschaal.



Grafiek 8: Beoordeling interactie

De gemiddelde scores liggen tussen 1,6 en 3,9, wat wijst op een overwegend neutrale tot gematigd positieve beoordeling, afhankelijk van het aspect.

De herkenning van de invoer is matig: alle groepen scoren laag ($M=2,3-2,5$) wat betekent dat deelnemers matig tevreden zijn met de invoerherkenning (oneens met de stelling). De subgroep die een geslaagde dialoog met SAM heeft gevoerd ($M=2,3$) scoort gelijk aan de controlegroep ($M=2,3$).

Respondenten in beide groepen geven aan dat de chatbot hen redelijk goed begreep. De gemiddelde scores variëren van 3,3 tot 3,6, wat duidt op een voldoende tot goed begrip. Ook hier scoort de controle groep iets hoger.

Beide chatbots komen vriendelijk over. De lage gemiddelde scores ($M=1,6-2,0$) zijn positief omdat lage waarden hier betekenen dat respondenten het helemaal niet eens zijn met de stelling (dus chatbot werd niet als onvriendelijk ervaren). De subgroep die een geslaagde dialoog met SAM heeft gevoerd, scoort op deze dimensie het best ($M=1,6$).

Verwelkoming bij het eerste gebruik verloopt voor beide chatbots goed. Gemiddelde scores ($M=3,7-3,9$), wijzen op een ruim voldoende tot goed gevoel van welkom. De subgroep die een geslaagde dialoog met SAM heeft gevoerd, scoort op deze dimensie het best ($M=3,9$).

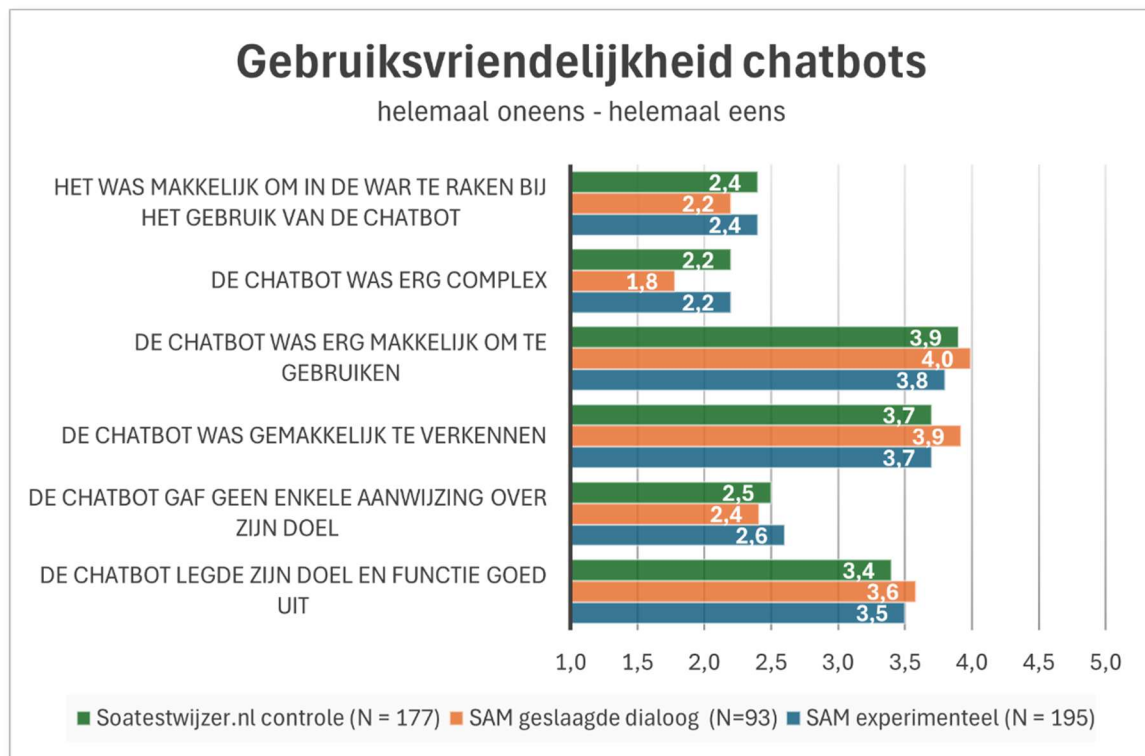
Er is een duidelijk verschil zichtbaar in de mate waarin de chatbots als robotachtig worden beoordeeld. SAM komt duidelijk robotachtiger ($M=3,3$) dan Soatestwijzer.nl die in de controlegroep is gebruikt ($M=2,9$).

De gemiddelde scores op robotachtigheid zijn in lijn met de beoordeling van de persoonlijkheid van de chatbot. Soatestwijzer.nl wordt voldoende als realistisch en boeiend ervaren ($M=3,2$) terwijl SAM met gemiddelde scores ($M=2,8-2,9$) als matig – net voldoende wordt beoordeeld.

Wat betreft de kwaliteit van de interactie met de chatbot kan samenvattend worden gesteld dat vriendelijkheid, verwelcoming en begrip positief worden beoordeeld. Ruimte voor verbetering voor SAM betreffen de natuurlijkheid (meer realistisch) meer vloeiende dialogen (nu te robotachtig en te weinig boeiend) en herkenning van de invoer.

4.3.3 Beoordeling gebruiksvriendelijkheid

In grafiek 9 zijn de gemiddelde scores op de gebruiksvriendelijkheid van de chatbots opgenomen.



Grafiek 9: Gebruiksvriendelijkheid chatbots

De gemiddelde scores op de aspecten die de gebruiksvriendelijkheid van beide chatbots meet zijn overwegend positief. Het gebruikersgemak wordt gemiddeld als ruim voldoende tot goed beoordeeld. Er zijn wel enkele aandachtspunten.

Gebruikers raakten niet makkelijk in de war tijdens het gebruik. De relatief lage scores voor SAM ($M=2,2$ - $2,4$) en Soatestwijzer.nl ($M=2,4$) zijn relatief positief omdat deelnemers het niet eens zijn met deze negatief geformuleerde stelling.

In lijn met de gemiddelde scores op de vorige stelling, worden de chatbots ook niet als complex ervaren. De gemiddeld lage score voor zowel SAM ($M=1,8$ - $2,2$) als in de controle arm ($M=2,2$) betekenen dat de chatbot niet als complex werden ervaren.

De chatbots werden ook beoordeeld als ‘makkelijk om te gebruiken’. De gemiddelde scores voor SAM ($M=3,8$ - $4,0$) en Soatestwijzer.nl ($M=3,9$) duiden op een ruim voldoende tot goed gebruiksgemak.

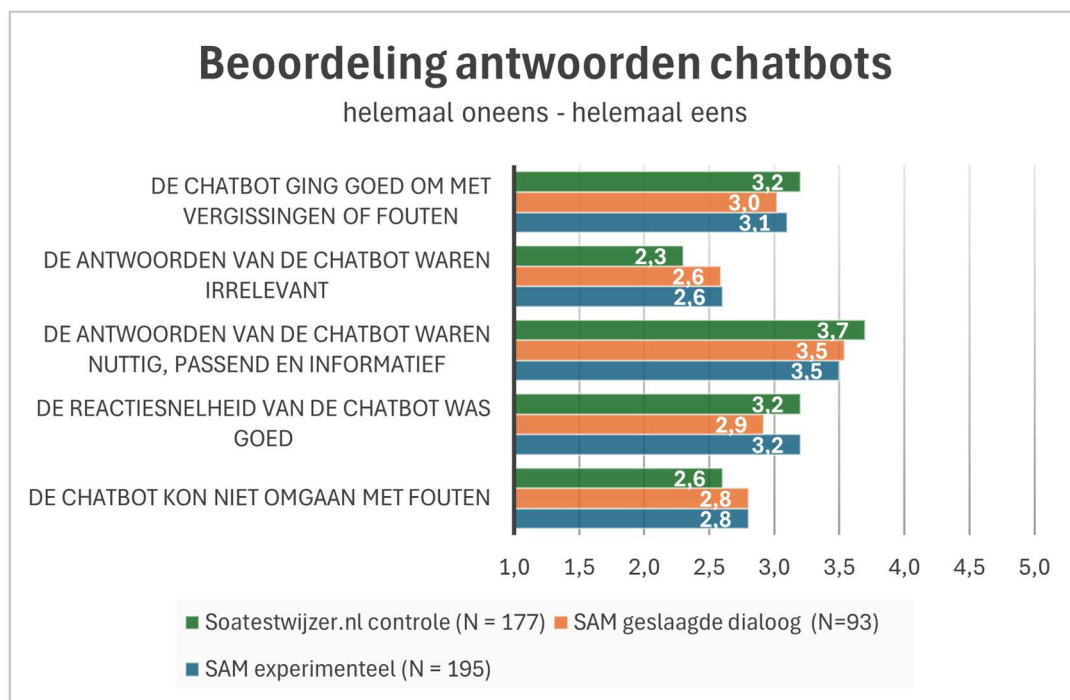
Verkenning van de chatbots is relatief makkelijk verlopen. Gemiddelde scores voor SAM ($M=3,7$ – $3,9$) en Soatestwijzer.nl ($M=3,7$) wijzen op een positieve ervaring bij het ontdekken van de mogelijkheden van de chatbots.

De gemiddelde scores op de twee stellingen die betrekking hebben op doel en functie van de chatbots laten ruimte voor verbetering zien. De stelling 'de chatbot gaf geen enkele aanwijzing over zijn doel' is beoordeeld met gemiddelde scores voor SAM (M=2,4 – 2,6) en voor de controle groep (M=2,5) wat betekent dat deelnemers niet sterk eens zijn met deze negatieve stelling, maar er is enige ruimte voor verbetering in duidelijkheid. De chatbots legde hun doel en functie relatief goed uit: met gemiddelde scores voor SAM (M=3,5–3,6) en Soatestwijzer.nl (M=3,4) wat ruim voldoende tot goed is , maar nog niet optimaal.

Samenvattend kan op basis van de gemiddelde scores op gebruiksvriendelijkheid worden gesteld dat beide chatbots ruim voldoende tot goed worden beoordeeld. Doel en mogelijkheden kunnen nog beter worden toegelicht. Opvallend is dat de subgroep die een geslaagde dialoog met SAM heeft gevoerd op alle dimensies van gebruiksvriendelijkheid gemiddeld (iets) beter scoort.

4.3.4 Beoordeling antwoorden

In grafiek 10 zijn de gemiddelde scores voor de beoordeling van de antwoorden van de chatbots opgenomen.



Grafiek 10: Beoordeling antwoorden

Beide chatbots lijken redelijk of voldoende om te gaan met fouten en vergissingen. De gemiddelde scores van SAM (M= 3,0 - 3,1) en Soatestwijzer.nl (M=3,2) op de bovenste stelling zijn in lijn met de negatief geformuleerde laatste stelling: SAM (M=2,8) en Soatestwijzer.nl (M= 2,6). Op deze dimensie lijkt voor beide chatbots verbetering op foutafhandeling mogelijk.

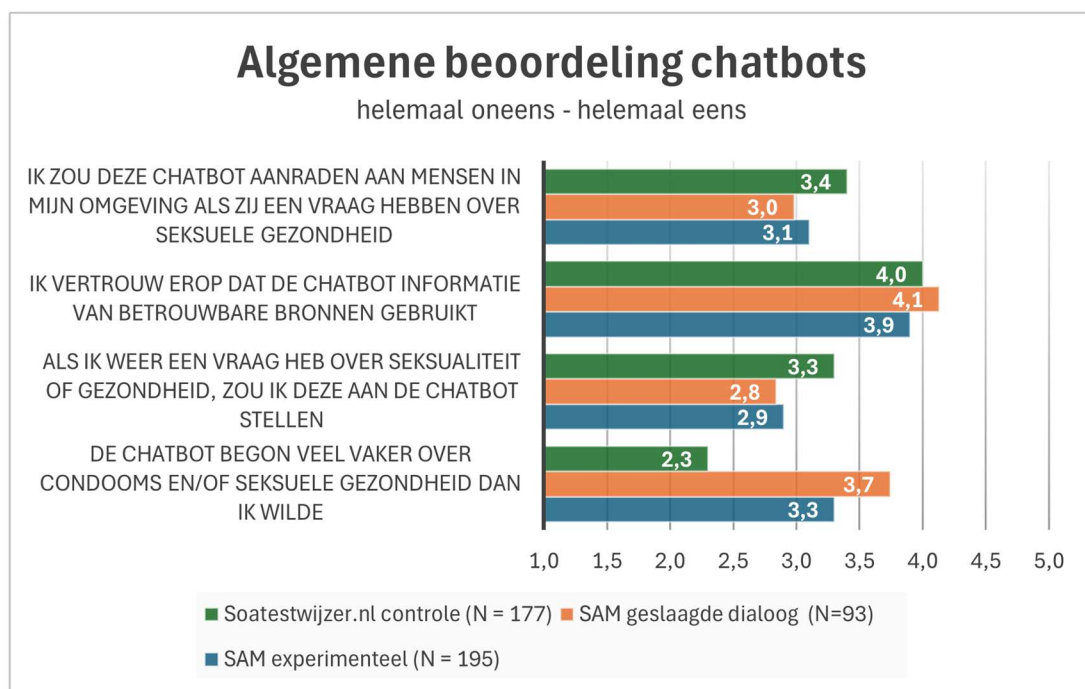
Wat betreft de beoordeling van de inhoud van de antwoorden is zichtbaar dat Soatestwijzer.nl beter wordt beoordeeld dan SAM. De gemiddelde scores van SAM op irrelevantie van de antwoorden (M=2,6) duidt op een matige tot voldoende beoordeling, de controle groep (M=2,3) oordeelde iets gunstiger. Beide chatbots scoren wel ruim voldoende tot goed op 'nuttig, passend en informatief' van de antwoorden met gemiddelde scores van SAM (M=3,5) en Soatestwijzer.nl (M=3,7).

De reactiesnelheid van de chatbots werden als voldoende - redelijk beoordeeld met gemiddelde scores van SAM ($M=2,9 - 3,2$) en Soatestwijzer.nl ($M=3,2$). Opvallend is dat de subgroep die een geslaagde dialoog met SAM heeft gevoerd op de waardering van de reactiesnelheid gemiddeld (iets) minder goed scoort ($M=2,9$).

Samenvattend kan worden gesteld dat wat betreft de beoordeling van de antwoorden van de chatbots de inhoud redelijk goed op orde is, maar wat betreft het omgaan met fouten en vergissingen en de reactiesnelheid verbeteringen mogelijk zijn.

4.3.5 Beoordeling antwoorden

In grafiek 11 zijn de gemiddelde scores op de algemene beoordeling van de chatbots opgenomen.



Grafiek 11: Algemene beoordeling chatbots

Hoewel SAM ($M=3,0 - 3,1$) een lagere gemiddelde score op 'aanraden aan mensen' in vergelijking tot Soatestwijzer.nl ($M=3,4$) is de gemiddelde score voor beide chatbots ruim voldoende. De subgroep die voldoende dialoog met SAM heeft gevoerd scoort lager ($M=3,0$).

Deze subgroep heeft er gemiddeld net iets meer vertrouwen ($M=4,1$) dat de informatie van betrouwbare bronnen gebruikt dan de gebruikers van Soatestwijzer.nl in de controle groep ($M=4,0$). Het vertrouwen in de bronnen van beide chatbots is op een goed niveau.

Als gebruikers weer een vraag zouden hebben over seksualiteit of gezondheid, is met name onder de gebruikers van SAM ($M=2,8 - 2,9$) lichte terughoudendheid waar te nemen in het voornemen weer de chatbot te gebruiken. Voor gebruikers van Soatestwijzer.nl ($M=3,3$) is de gemiddelde score beter.

In lijn met de opdracht aan SAM om de dialogen met gebruikers proactief te richten op condooms, is de gemiddelde score in de subgroep die voldoende dialoog met SAM heeft gevoerd op het item 'vaker over condooms en/of seksuele gezondheid beginnen dan ik wilde' veel hoger ($M=3,7$) dan in de controlegroep ($M=2,3$).

Samenvattend kan op basis van de gemiddelde scores van de algemene beoordeling worden gesteld dat er behoorlijk hoog vertrouwen is in de gebruikte informatiebronnen. Dat is een mooie basis voor hergebruik van beide chatbots, maar zowel de mate waarin men anderen aanbeveelt ook gebruik te maken van de chatbot als persoonlijk herhalingsgebruik zijn nog voor verbetering vatbaar. Heel zichtbaar is dat SAM gebruikers meer stuurt naar dialogen over condooms en seksuele gezondheid dan Soatestwijzer.nl.

4.3.5 Sterke punten en suggesties voor verbetering volgens gebruikers

In de vragenlijst is een tweetal open vragen opgenomen die relevant zijn voor de doorontwikkeling van de chatbots:

- ✓ Wat vindt u goed aan de chatbot?
- ✓ Wat zou volgens u verbeterd moeten worden aan de chatbot?

De antwoorden van de subgroep die voldoende dialoog hebben gevoerd met SAM (N=93) op deze twee vragen zijn in paragraaf 4.3.5.1 samengevat.

4.3.5.1 Wat vinden gebruikers goed aan SAM?

Van deze 93 respondenten vonden zes deelnemers niks goed aan SAM.

Duidelijkheid & kwalitatieve antwoorden

De gebruikers ervaren de antwoorden van de chatbot als helder, gestructureerd en goed te begrijpen. De lengte van de antwoorden vinden ze passend; niet te kort en niet te lang, maar uitgebreid genoeg en toch 'to the point'. SAM pakt ook complexere vragen op en kan deze ook goed uitleggen.

Bronnen & extra informatie

Een belangrijk pluspunt dat vaak genoemd wordt, is dat SAM standaard doorverwijst naar betrouwbare bronnen en links toevoegt. Dit geeft de gebruikers de mogelijkheid om verder te lezen, zich te verdiepen in een onderwerp en meer informatie te krijgen. Het vermelden van bronnen maakt dat de antwoorden nuttig en geloofwaardiger worden gezien.

Snelheid & gebruiksgemak

De chatbot reageert snel en direct, zonder dat de gebruikers zelf uitgebreid moeten zoeken naar informatie en dit wordt erg gewaardeerd. Daarnaast wordt ook benadrukt dat het gebruik van de chatbot laagdrempelig en anoniem is, het maakt het makkelijker en fijner om vragen te stellen, ook als deze over gevoelige onderwerpen gaan en zorgt ervoor dat hier minder barrières ervaren worden.

Persoonlijk & interactie

De chatbot komt vriendelijk over en geeft de gebruikers het gevoel dat de bot echt naar ze luistert. Daarnaast heeft de SAM een open houding en geen veroordelende toon. Het stellen van vervolgvragen wordt gezien als een manier om het gesprek actief en levendig te houden en om verder te helpen, ook als de gebruiker even niet weet wat ze precies willen vragen. De toon wordt omschreven als positief, geruststellend, persoonlijk en empatisch.

Betrouwbaarheid & veiligheid

Gebruikers vinden het prettig en belangrijk dat SAM wijst op veiligheid en het aangeven van grenzen. De chatbot geeft betrouwbare en realistische informatie, en het voelt vertrouwd en veilig om een gesprek te voeren. Het laten zien van een betrouwbare organisatie die achter de chatbot zit, draagt ook bij aan het gevoel van vertrouwen en veiligheid.

4.3.5.2 Wat zouden gebruikers verbeteren aan SAM

Van de 93 respondenten hadden twee deelnemers geen verbeterpunten voor SAM.

Reactiesnelheid

Veel gebruikers vinden dat de reactiesnelheid verder verbeterd kan worden, nu lijkt het vaak alsof de chatbot te lang moet 'nadenken'. Dit vertraagt het gesprek en kan ertoe leiden dat ze afhaken.

Herhaling & relevantie

Een terugkerend verbeterpunt is dat SAM meer variatie mag toevoegen in de antwoorden, weetjes of vervolgvragen. Gebrek aan variatie maakt het gesprek wat eentonig en repetitief, waardoor het ook meer het gevoel van een robot geeft. De vervolgvragen sluiten soms ook niet aan op het onderwerp dat de gebruiker aansneed of gaf irrelevante weetjes. Daarnaast zijn de suggesties of vervolgvragen van SAM soms te sturend of 'te standaard' waardoor het dwingend en onpersoonlijk kon overkomen.

Inhoud & diepgang

Gebruikers vinden dat er soms wat meer diepgang in de antwoorden mag zitten en beter aansluiten op de specifieke vragen. De antwoorden lijken nu af en toe wat te algemeen of oppervlakkig. Een korte toelichting zouden ze ook waarderen bij de bronnen, zodat duidelijk is wat deze toevoegen.

Bronnen & betrouwbaarheid

Hoewel de bronnen zeer gewaardeerd werden, vinden gebruikers dat ze niet altijd de meest relevante bronnen krijgen. Hier zou specifieker naar gekeken moeten worden zodat ze beter aansluiten op de vragen en behoefte van de gebruikers. Het zou ook kunnen helpen om meer betrouwbare bronnen toe te voegen, denk hierbij aan thuisarts.nl.

4.3.6 Conclusies usability chatbots

De beoordeling van de chatbots laat een overwegend positief beeld zien van gebruiksvriendelijkheid, interactie en betrouwbaarheid van informatie. Gebruikers ervaren beide chatbots als vriendelijk, voldoende begrijpelijk en eenvoudig te gebruiken, met hoog vertrouwen in de gebruikte bronnen. Een belangrijk pluspunt dat vaak genoemd wordt, is dat SAM standaard doorverwijst naar betrouwbare bronnen en links toevoegt. Tegelijkertijd wijzen zowel kwantitatieve scores als kwalitatieve opmerkingen op verbeterpunten: SAM wordt als te robotachtig ervaren, foutafhandeling en reactiesnelheid kunnen beter, en de uitleg van doel en mogelijkheden verdient meer aandacht. Daarnaast blijkt uit open antwoorden dat gebruikers van SAM authenticiteit en variatie belangrijk vinden voor een prettige ervaring. Aanbeveling aan anderen en hergebruik blijven enigszins achter, wat suggereert dat verdere optimalisatie van dialoogkwaliteit en personalisatie nodig is om acceptatie en herhaalgebruik te stimuleren.

BIJLAGE 1: OPDRACHTOMSCHRIJVINGEN

Experimentele arm

Fijn dat u de chatbot wil gaan testen. Nadat u de chatbot getest heeft, kunt u weer teruggaan naar dit venster waar de vragenlijst staat. Is het venster met de vragenlijst gesloten of verlopen? Geen probleem – open dan de vragenlijst opnieuw via de link in de uitnodigingsmail. U kunt dan verder gaan waar u gebleven was.

De chatbot heet Sam en is speciaal ontwikkeld om vragen over relaties en seksualiteit te bespreken. Met dit onderzoek willen we graag achterhalen hoe u Sam beoordeelt. U mag Sam uitproberen. Vanwege de gevoeligheid van het onderwerp, worden alle gesprekken die met Sam worden gevoerd vertrouwelijk en anoniem behandeld.

- ✓ We willen u het volgende vragen:
- ✓ Lees eerst alle instructies op dit scherm voordat u naar de chatbot gaat.
- ✓ De link naar de chatbot staat onderaan op dit scherm.
- ✓ Ga met Sam een gesprek aan. We willen u vragen het gesprek in 1 keer voeren en niet tussentijds te onderbreken.
- ✓ Start het gesprek met een vraag over relaties of seksualiteit. Dat kan van alles zijn, begin vooral met een vraag die u zelf interessant vindt of waar u benieuwd naar bent. Bijvoorbeeld over versieren, daten of standjes.
- ✓ Voer een gesprek van ongeveer 6 – 7 minuten.
- ✓ Stel daarna een vraag over veilig vrijen, bijvoorbeeld over condooms of soa's.
- ✓ Stop na totaal 10 minuten met het gesprek met Sam.
- ✓ Ga daarna terug naar de vragenlijst en vul het verder in.

Controle arm

Fijn dat u meedoet aan ons onderzoek in opdracht van Soa Aids Nederland.

We willen u eerst een aantal vragen over seksuele gezondheid stellen. Daarna vragen we u om de chatbot van Soa Aids Nederland te testen. Tot slot vragen we hoe u het gebruik van de chatbot heeft ervaren.

Voor het testen van de chatbot wordt u doorverwezen naar een externe website van Soa Aids Nederland. Daar gelden de privacyregels van Soa Aids Nederland. Soa Aids Nederland verwerkt het gesprek dat u heeft met de chatbot anoniem. Daarna kunt u de vragenlijst bij ons vervolgen.

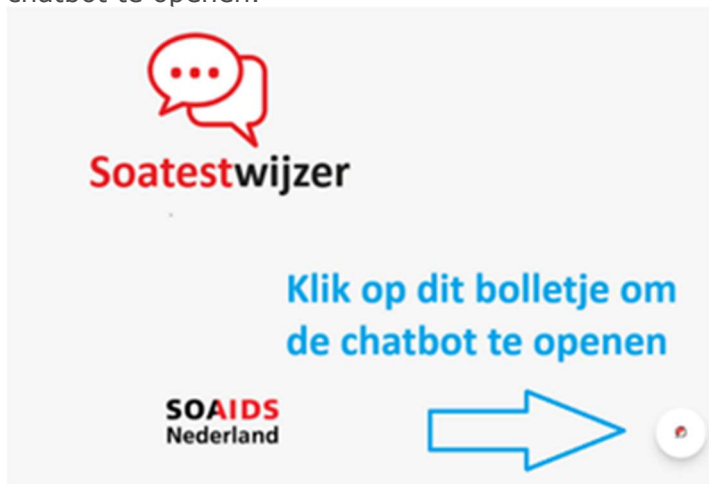
Fijn dat u de chatbot wil gaan testen. Nadat u de chatbot getest heeft, kunt u weer teruggaan naar dit venster waar de vragenlijst staat. Is het venster met de vragenlijst gesloten of verlopen? Geen probleem – open dan de vragenlijst opnieuw via de link in de uitnodigingsmail. U kunt dan verder gaan waar u gebleven was.

De komende 10 minuten vragen we u de chatbot te testen die Soa Aids Nederland heeft gemaakt. De chatbot heet Sanne. Sanne helpt mensen op weg, die vragen hebben over soa's, hiv, testen, behandeling enzovoorts. U helpt Soa Aids Nederland de chatbot beter te maken door vragen in uw eigen woorden te stellen.

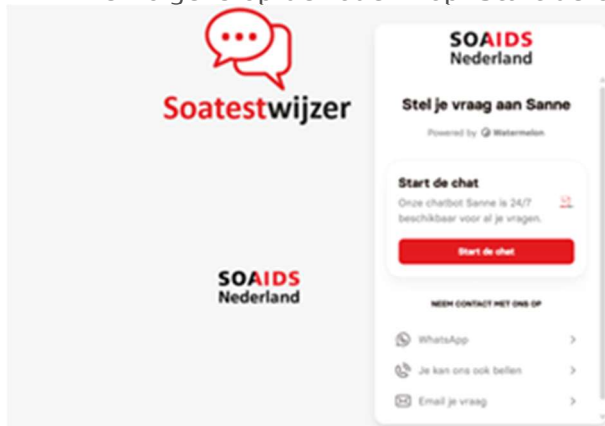
We willen u het volgende vragen:

- Lees eerst alle instructies op dit scherm voordat u naar de chatbot gaat.
- De link naar de chatbot staat onderaan op dit scherm.

- Nadat u op de link geklikt heeft, moet u op het bolletje rechts onderin klikken om de chatbot te openen.



- Klik vervolgens op de rode knop 'Start de chat'.



- Wat zou u willen weten, over soa's of soa-testen? Stel uw vraag aan de chatbot in uw eigen woorden. Weet u zelf geen vraag? Stel dan een vraag vanuit één van de onderstaande verhalen van Jake, Roel of Holly.

- Lees de reactie van de chatbot.

- Vindt u het antwoord niet duidelijk, dan kunt u meer vragen stellen.

- U kunt op elk antwoord klikken en aangeven wat u van het antwoord vindt. Dat doet u door een duimpje omhoog of omlaag te geven.

- U bepaalt zelf welke vragen u wel of niet wilt stellen. We verzoeken u 10 minuten te oefenen. Daarna kunt u de chatbot afsluiten en teruggaan naar deze vragenlijst.

Verhalen* om mee te oefenen

Jake: Jake en zijn vriendin zijn niet meer samen. Jake maakte het uit, omdat zij met iemand anders seks had gehad. Zijn ex vertelde hem dit een half jaar nadat het gebeurd was. Zijn ex vertelde dat zij seks met condoom had en geen klachten heeft gehad. Jake heeft in het afgelopen jaar ook geen klachten gehad. Jake wil toch weten of hij zich moet laten testen en zo ja, welke soa's dan.

Roel: Roel voelt zich al een week ziek. Het voelt als een flinke griep. Hij heeft koorts en zweet 's nachts veel. Overdag voelt hij zich niet fit. Een paar weken geleden had Roel seks zonder condoom. Hij maakt zich zorgen of hij misschien hiv heeft opgelopen. Hij vindt het lastig om naar zijn huisarts te gaan. Roel wil eerst meer weten of hij misschien hiv kan hebben. Kan hij ook zelf een test kopen om op hiv te testen?

Holly: Holly heeft een vaste relatie. Sinds een paar weken heeft ze last van kleine bultjes bij haar geslachtsdelen. Holly is bang dat het genitale wratten zijn. Holly en haar huidige partner zijn al twee jaar samen en hebben alleen met elkaar seks. Voordat Holly haar partner leerde kennen, had Holly met twee andere mensen seks gehad. Holly wil weten welke klachten bij genitale wratten passen. Vroeger had ze nergens last van, waarom nu dan wel? Ook dacht Holly tegen genitale

wratten beschermd te zijn, omdat ze als tiener gevaccineerd is tegen HPV.

*Deze verhalen zijn niet echt. Als u iemand in het verhaal herkent, is dat toeval!

Heeft u bovenstaande instructies gelezen? Dan kunt u de chatbot gaan testen door naar [Soa Aids Nederland – Testpagina](#) te gaan. Door [hier](#) te klikken, opent de chatbot in een nieuw venster.

Als u klaar bent met het testen van de chatbot, klik dan op 'volgende vraag'.