



aanpak 
**hegeleidings
ethiek**

Verslag workshop
Aanpak Begeleidingsethiek
**Digitale avatars in
communicatie met burgers**
Nationale Politie

ECP

Platform voor de
InformatieSamenleving

Voorwoord

De toenemende digitalisering brengt ingrijpende veranderingen met zich mee, in allerlei sectoren. Nieuwe mogelijkheden voor communicatie, monitoring en analyse roepen vragen op. Wat is de positie van de mens in deze omgeving, hoe staat het met data, met privacy, wat doen algoritmes en wat willen gebruikers eigenlijk? Begeleidingsethiek is ethiek die zich specifiek met dit soort vragen bezighoudt. Hij is gebaseerd op techniekfilosofie, die uitgaat van eeuwenoude verwevenheid tussen mens en technologie.

De Aanpak Begeleidingsethiek is een concrete aanpak waarin betrokkenen met elkaar in dialoog gaan over de effecten van de nieuwe technologie én de waarden die daarbij in het geding komen. Dat zijn vaak waarden gelieerd aan autonomie van de gebruiker, efficiëntie van het proces, transparantie van het algoritme, privacy, et cetera.

Tijdens de sessie komen verschillende stakeholders als gebruikers, ontwikkelaars, beleidsvormers en beslissers met elkaar in gesprek. Na de dialoog hebben de deelnemers ethische handelingsopties gegenereerd, waarvan verschillende vaak direct opgepakt kunnen worden. Er is gezamenlijk gekeken welke waarden we in het digitale domein belangrijk vinden en hoe we die willen verankeren en borgen in digitale processen en handelingen. Dit alles om ook bij verdere digitalisering op het vertrouwen van de samenleving en de participanten kunnen blijven rekenen. Dit is niet in één stap te realiseren het is een continu proces waarin deze workshop een schakel is.



Workshop aanpak begeleidingsethiek

Casusaandragers: Marian Luursema en Manon van Dunnen (Nationale Politie)

Moderatoren: Aldert de Jongste en Lianne Kramer (co-moderator)

Op 2 oktober 2023 organiseerde ECP op verzoek van het ministerie van Binnenlandse Zaken en Koninkrijksrelaties een sessie begeleidingsethiek over de casus 'de inzet van digitale avatars in communicatie met burgers'. Deze sessie maakt onderdeel uit van de BZK reeks Waardengedreven Dialogen. Aan de workshop namen 14 deelnemers (zie bijlage voor specificatie) deel onder leiding van Aldert de Jongste (ECP) en Lianne Kramer (Hooghiemstra & Partners).

De doelstelling is tweeledig:

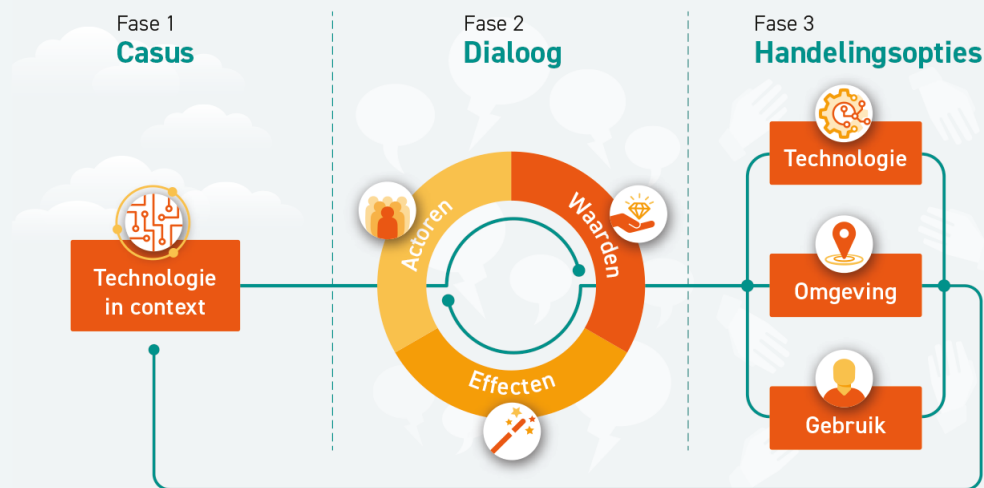
- ▷ Leren tijdens de workshop
 - Welke effecten, waarden en actoren worden genoemd
 - Welke handelingsopties zijn er om de toepassing te verbeteren
- ▷ Leren over de Aanpak Begeleidingsethiek
 - Hoe kan de organisatie deze methodiek gebruiken rondom ethische kwesties en technologie-implementaties

Het verslag bevat de volgende elementen:

- ▷ Toelichting op de Aanpak Begeleidingsethiek
- ▷ Weerslag van de workshop



Aanpak begeleidingsethiek



De aanpak bevat de volgende fasen:

Fase 0 Introductie

Introductie over de doelstelling en een toelichting op het model en het gedachtengoed

Fase 1 Toelichting

Hoe ziet de technologie eruit en in welke context wordt deze toegepast

Fase 2 Dialoog

- ▷ Een korte ronde waarin de deelnemers aan de workshop de betrokken actoren benoemen
- ▷ Brainwrite waar deelnemers mogelijke effecten benoemen en bespreken
- ▷ Benoemen van waarden die een rol spelen bij die effecten

Fase 3 Handelingsopties

In subgroepjes gaan de deelnemers op zoek naar handelingsopties vanuit de technologie, de omgeving en het individu.



Fase 1

Technologie en context

Uitleg over de toepassing/technologie

De aanwezige organisaties willen de mogelijke inzet van een online digitale persona voor contact met burgers onderzoeken. Inspiratie bij de sessie is de huidige chatbot Wout van de Nationale Politie. De inzet van deze technologie is nog zeer speculatief. Voor de sessie wordt ervan uitgegaan dat de avatar wordt ingezet door de politie voor “eerstelijns contact” met burgers. Primaire rol van de avatar is basis-vragen beantwoorden en doorverwijzen naar relevante afdelingen. Centraal in deze sessie staat de verschijningsvorm, het uiterlijk, van de avatar. We gaan uit van een niet van een persoon te onderscheiden bewegende beeltenis. De audio-opties, het script en de content van het gesprek laten we buiten beschouwing.



Fase 2

Dialogo

In deze tweede fase gaan de deelnemers in gesprek over wie er betrokken zijn bij de inzet en het gebruik van de avatar. Ook buigen zij zich over de positieve en negatieve effecten van deze inzet, en benoemen zij belangrijke waarden waar rekening mee gehouden moet worden bij de inzet van een dergelijke toepassing.

Actoren:

Bij de actoren is de vraag wie er betrokken is of geraakt wordt door de case. De deelnemers aan deze specifieke sessie is minder divers als vaak voor een sessie begeleidingsethiek. Voor deze workshop is gekozen om vertegenwoordigers van een aantal verschillende overheidsorganisaties aan te laten sluiten. Ze noemen de volgende betrokkenen:

- Burgers
- Medewerkers organisatie die avatar inzet
- Websitebouwers
- Politie op straat
- Management
- Redacteuren/contentmakers
- Tweede lijn
- Deskundigen
- Maatschappelijk middenveld
- Ketenpartners, zoals gemeenten, organisaties in de gezondheidszorg
- UX-designers
- Data-analisten
- Psychologen
- Media
- Ondernemingsraad



De vraag aan de deelnemers was om in het vervolg van de sessie ook te proberen vanuit deze perspectieven hun inbreng te geven.



Effecten:

De vraag is hier welke positieve en negatieve effecten de invoering en het gebruik van de avatar hebben.

Positieve effecten:

- Gevoel vd gebruiker erbij te horen, herkenbaarheid van de menselijke gedaante
- Door menselijk karakter avatar te zien, kun je meer emotie/betrokkenheid uitstralen en uitnodigen hart uit te storten
- Laagdrempelig voor mensen met beperking (bv. Zien en horen spreken)
- Gevoel van tijd en aandacht, doordat het aanvoelt als een echt persoon
- Je hoeft niet heel digivaardig te zijn omdat je spraakavatar gebruikt
- Mensen vertrouwen sneller op antwoord als je met avatar praat
- Bespaart personeelskosten
- Bot is niet oordelend/neutraal
- Je kunt de bot herkenbaar maken voor specifieke doelgroep
- Uiterlijk geeft lading mee, ethisch/geslacht/religie - minder neutraal dus dan een (sec) textbot. Kan effect hebben op communicatie, meer vertellen of juist minder.
- Digitale karakter maakt het proces en verschijningsvorm makkelijk aanpasbaar

Negatieve effecten:

- Extra confrontatie echte wereld: bot is mogelijk ideale versie, neutraal, inlevend, etc. Echte medemensen/medewerkers dragen bagage mee.
- Chatwens is gedeeltelijk juist afstandelijkheid (anonieme tekstinterface), door chatbot in menselijke avatar te veranderen verliest het mogelijk laagdrempeligheid
- Focus op digitaal gaat (vaak) ten koste van fysieke omgeving
- Risico op stereotypering bij uiterlijk geven aan avatar
- Mensen kunnen denken dat het echte persoon is, terwijl dat niet zo is (illusie van echt persoon)
- Je denkt menselijke maat-antwoord te krijgen, maar dat is het in feite niet. Er wordt alleen een proces gevolgd.
- Vervreemding door gescipte avatar.tenzij mensen niet echt weten wat er aan de hand is - dan denken ze invloed te kunnen uitoefenen. Maar dat is dan een illusie, wat vinden we daarvan?
- Pos/neg: Emotionele lading brengt verantwoordelijkheid met zich (bij 113 bijvoorbeeld). Antwoord wordt anders gelezen als het een menselijk antwoord is.



- Door de keuze voor een bepaald uiterlijk kan het voelen als manipulatie
- Tegenstrijdigheid in het systeem, straal uit mens, maar is een bot
- Met de bot bepaal je een norm, bijvoorbeeld een neutrale uitstraling
- Onduidelijk wie meekijkt, wat gebeurt er met input - dat bepaalt het systeem
- Beeldvorming dat politie in fysieke wereld terugtrekkende beweging maakt
- Kan tot nieuwe stereotypering leiden door avatar in te stellen op bepaald geslacht/huidskleur, hoe denken mensen over mensen met bepaalde kenmerken doordat avatar (on)wenselijk antwoord geeft?
- Excuus niet te diversificeren binnen de organisatie, want de avatars zijn toch divers.
- Doet afbreuk aan menselijk contact met overheid

Waarden:

De relevante waarden zijn door de facilitators geïnventariseerd terwijl de deelnemers bezig waren met het benoemen van effecten. Achter veel effecten gaan waarden schuil. Die lijst is gedeeld met de deelnemers en door hen besproken en aangevuld. Dat leverde de volgende lijst met waarden op:

Emotie
Menselijkheid
Authenticiteit
Kwaliteit
Herkenbaarheid
Inclusiviteit
Toegankelijkheid
Verbondenheid
Autoriteit
Geloofwaardigheid
Efficiëntie
Juistheid
Gelijkheid
Uniformiteit
Vertrouwen
Integriteit
Flexibiliteit



Verantwoordelijkheid
Privacy
Zichtbaarheid

We hebben de volgende drie waarden geselecteerd voor fase 3:

- **Authenticiteit**
- **Verantwoordelijkheid**
- **Inclusie.**

Fase 3

Handelingsopties

In het proces van het opstellen van de handelingsopties zijn de waarden uit de vorige fase meegenomen. De deelnemers kregen een uitleg over de verschillende categorieën binnen de handelingsopties. Het kan bijvoorbeeld gaan over het ontwerp van de technologie, het aanpassen van de omgeving waarbinnen de toepassing functioneert en het gedrag van mensen. De deelnemers worden in drie werkgroepen verdeeld en gaan ieder aan de slag met een van deze drie gespreksonderwerpen. Dat leidt tot handelingsopties per categorie.

Technologie:

Authenticiteit

Aandachtspunt is dat er nog veel waarde gedreven vragen ervoor liggen die niet beantwoord zijn net als waarom, met wel doel en proportionaliteit.

1. Niet doen
2. Nepper maken (cartoonesk) en duidelijk communiceren dat je een bot bent en wat je limieten zijn.
3. Perfecter maken zodat mensen niet doorhebben dat je nep bent, moet wel passen bij doel en hoort waardenvol gesprek aan vooraf te gaan omdat je tegen allerlei waarden ingaat.

Verantwoordelijkheid

Eigenlijk een paraplueterm voor allerlei waarden die je als organisatie moet meenemen in ontwerp en implementatie, denk aan beschikbaarheid (wanneer doe je onderhoud), integriteit, begrijpelijkheid, gelijkwaardigheid,...etc.

Inclusie



Dit roept een aantal vragen op zoals:

1. Wie bepaalt voor de ander hoe diegene benaderd wil worden
2. Welke waarden laat je prevaleren; geef je de ander waar deze behoefte aan heeft of daag je degene uit om buiten de comfortzone te komen. Dit hangt ook samen met wat je als verantwoordelijkheid ziet voor je organisatie en waar je voor staat.

Keuze van wat je als avatar aanbiedt is niet neutraal, bied je 1 type, laat je mensen kiezen, of genereer je iets ad random. Stel je hebt te maken met een extreem racistisch persoon, wat je kiest sluit andere opties uit en bepaalt mede waar je (voor) staat.

Omgeving:

Authenticiteit

- Invoeren overheidskeurmerk
- Informeren van de burger over wat een avatar is en wat die wel en niet kan betekenen

Verantwoordelijkheid

- Stevig proces bouwen voor invoering avatar
- Testen van de avatar als alternatief t.o.v. een mens
- Inrichten goede governance (wie is waarvoor verantwoordelijk?)
- Doorverwijzingen naar andere organisaties als dat nodig is, liefst binnen één loket van avatars

Inclusie

- Gemiddelde persoon genereren o.b.v. medewerkers bij de organisatie om uiterlijk avatar te bepalen (maar dit kan juist ook ongewenst effect hebben dat als de organisatie geen goede afspiegeling is van de samenleving, dit beeld in de avatar naar voren komt)
- Juist een random avatar weergeven
- Avatar moet WCAG-proof zijn

Mens:

Authenticiteit

- Duidelijk maken aan gebruikers dat het een Avatar betreft
- Duidelijk maken aan gebruikers of/dat mogelijk verschillende Avatars gekozen of gebruikt worden
- Avatar moet voor gebruikers duidelijk de organisatie uitstralen (politie), dus



dit ook duidelijk communiceren

- Zekerheid qua herkenbaarheid, gebruikers dus ook duidelijk instrueren hoe zij de authenticiteit van een bot kunnen verifiëren

Verantwoordelijkheid

- Bot moet agitatie herkennen (technologie)
- Bot moet feitelijk blijven niet gaan speculeren op basis van de agitatie van de burger(technologie)
- Geografische indeling in systeem zodat bij escalatie direct regionaal/lokaal gereageerd kan worden off-line. Zowel eindgebruikers als ondersteuning hierin meenemen.
- Bewustzijn creëren bij gebruikers dat bot feilbaar is, in tegenwicht op menselijke trek om computers als foutloos te zien.

Inclusie

- Bewustzijn creëren dat je verschillende bots kan tegenkomen, dus ook van verschillende etniciteiten.



Terugblik en afronding

Aan het einde van de workshop wordt teruggekeken en besproken wat de Aanpak Begeleidingsethiek concreet heeft opgeleverd.

De deelnemers geven aan dat het veel nieuwe inzichten oplevert en dat het leuk is om veel verschillende perspectieven te horen. Tegelijkertijd wordt aangegeven dat niet alle expertise van iedereen goed aan bod heeft kunnen komen. De casus zelf was wellicht wat erg 'next level' voor sommige organisaties. Mede daardoor was de casus ook niet helemaal scherp neer te zetten. Een concretere casus zou de sessie nog beter hebben gemaakt. Omdat de workshop over een relatief speculatieve casus ging misten sommige aanwezigen de voorvraag 'waarom willen we deze technologie inzetten?'. De combinatie van voorziene negatieve effecten en de moeilijkheid om deze met handelingsopties te mitigeren was voor sommige aanwezigen reden om te stellen dat invoer geen goed idee is.

Al met al was het een leerzame sessie, met mensen uit verschillende organisaties waartussen volgens enkelen meer contact zou mogen bestaan rondom soortgelijke onderwerpen.



De deelnemers

1. Manon van Dunnen	Politie
2. Marian Luursema	Politie
3. Barbara van Berkel	Politie
4. John Bloebaum	Politie
5. Casper Doing	Politie
6. Jeroen Marttin	Studio Winegum
7. Merijn Spelt	Studio Winegum
8. Ingrid Zondervan	Ministerie van Binnenlandse Zaken
9. Wim van Melzen	Gemeente Zoetermeer
10. Gerald Munters	Belastingdienst
11. Mark Schoondorp	Belastingdienst
12. Jan Sluyterman	Rathenau Instituut
13. Marit de Jong	Rathenau Instituut
14. Harry van Geijn	Microsoft

Moderatoren:

Aldert de Jongste (ECP), Lianne Kramer (Hooghiemstra & Partners)



aanpak 
**begeleidings
ethiek**



Platform voor de
InformatieSamenleving

Verslag workshop
Aanpak Begeleidingsethiek
Digitale avatars in
communicatie met burgers

Meer informatie over de Aanpak Begeleidingsethiek,
waaronder dit verslag, vindt u op

www.begeleidingsethiek.nl