



Platform 
**Begeleidings
ethiek**

Verslag workshop
Aanpak Begeleidingsethiek
**Generatieve AI voor creatief
promotioneel materiaal**

RTL



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

Initiatiefnemers:

- Daan Odijk (RTL)

Moderatoren:

- Edwin Borst (ECP | Platform voor de InformatieSamenleving)
- Daniël Tijink (ECP | Platform voor de InformatieSamenleving)

Op 12 december 2023 organiseerde RTL op initiatief van het ministerie van Binnenlandse Zaken een sessie begeleidingsethiek over de inzet van generatieve AI voor creatief promotioneel materiaal. Aan de workshop namen 14 deelnemers (zie bijlage voor specificatie) deel.

De doelstelling is tweeledig:

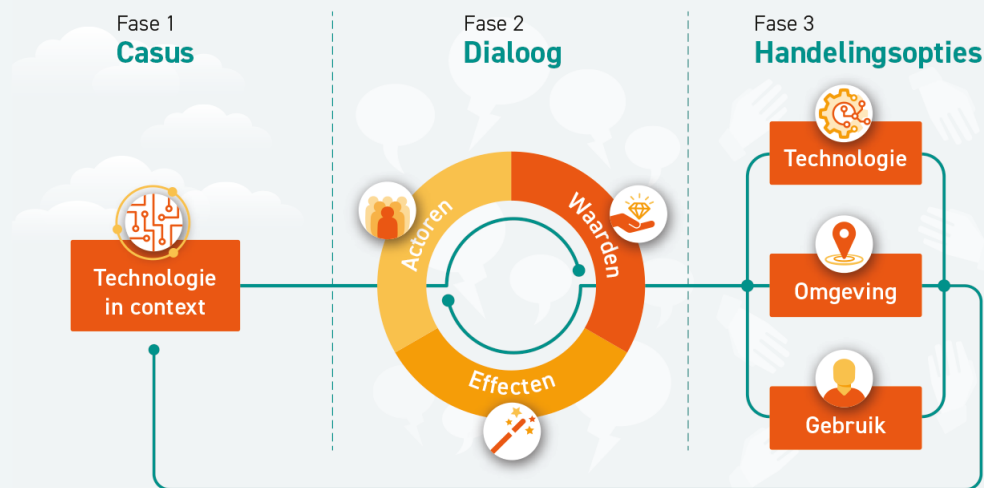
- ▷ Leren tijdens de workshop
- Welke effecten, waarden en actoren worden genoemd
- Welke handelingsopties zijn er om de toepassing te verbeteren
- ▷ Leren over de Aanpak Begeleidingsethiek
- Hoe kan de organisatie deze methodiek gebruiken rondom ethische kwesties en technologie-implementaties

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

De sessie vindt plaats bij RTL in Hilversum. De specifieke technologie en context zijn uitgelegd door Marco Gijsen. Hij werkt bij Ad Alliance, dat zich toelegt op het maken van maatwerkcampagnes. Eén van de dingen die ze maken zijn de billboards: beelden vlak voordat een televisieprogramma begint, vaak gepaard gaand met de tekst: “dit programma wordt mede mogelijk gemaakt door...” Voor zo’n billboard zijn foto’s nodig die specifiek toegesneden zijn op enerzijds het te verkopen product en anderzijds het programma waarvoor het komt. Dat betekent in de praktijk een shoot op locatie met modellen. De casus gaat over de inzet van generatieve AI (Midjourney) in deze context. Een voorbeeld is een billboard dat Rivella voor het programma B&B Vol Liefde wilde laten maken. Via Midjourney werden verschillende foto’s ‘gemaakt’. De stellen bestaan niet echt, maar zijn samengesteld uit de totale hoeveelheid beelden die er zijn. Het was makkelijk om te variëren, bijvoorbeeld een ouder stel of een jonger stel, verschillende settings. Het leverde snel bruikbare resultaten op. Wel zijn er nog andere programma’s nodig, zoals Firefly van Adobe, om het af te maken. Flesjes Rivella moeten worden ingeplakt, want die kwamen niet naar voren. Ook zijn vingers vaak niet goed. Om te voorkomen dat het onverhoopt toch een gezicht van een bestaand iemand werd, is reverse image search toegepast. Het beeld werd in image search van google gestopt en dan werd snel duidelijk of er nauwe overeenkomsten waren. In dat geval werd een ander beeld gebruikt. Een ander voorbeeld ging om een billboard voor Make-a-wish, waarin zieke kinderen figureerden. Een meisje had buisjes in haar neus, en een afbeelding ervan was niet te genereren in Midjourney. Hier is de achtergrond wel met Midjourney gemaakt, maar is de foto van een deelnemer aan Make-a-wish gebruikt.

De technologie zou je je kunnen voorstellen als twee robots. Robot 1 weet welke woorden bij welk plaatje horen. Robot 2 weet hoe hij van een bepaalde beschrijving een mooi plaatje kan maken. Je zou dat kunnen vergelijken met een goede schilder, die kan ook uit zijn hoofd een schilderij maken van bijvoorbeeld een mens die aan tafel zit. In het hoofd van de schilder zitten veel beelden van mensen aan tafels en daar wordt een schilderij uit gecomponeerd.



Fase 2

Dialogo

In deze tweede fase gaan de deelnemers in gesprek over wie er betrokken zijn bij de inzet en het gebruik van generatieve AI voor creatief promotioneel materiaal. 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 tafel vertegenwoordigen al een deel van die actoren. Ze noemen de volgende betrokkenen:

- Fotografen
- Opdrachtgever van reclame (zoals hier bijv. Rivella)
- Juristen (van RTL, van de fotograaf etc.)
- Publiek/kijkers
- Gezichtsleverancier
- Prompter
- Vormgever
- Creatieve bureaus
- Midjourney (en andere GAI)
- Adobe Firefly/Photoshop (en andere beeldbewerkingsprogramma's)
- Voice-overleveranciers
- Creatives
- Directie RTL
- Toezichthouder: Reclame Code Commissie, Commissariaat van de Media
- BZK
- OCW
- Data scientists RTL
- Platform, de zender
- (Videoland/TV)
- Modellen(bureau's)
- Aandeelhouders
- NGO's, vertegenwoordigers van specifieke groepen
- Social media (denk aan TikTok, Facebook)

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 generatieve AI voor creatief promotioneel materiaal hebben.

Positieve effecten

- Draagt bij aan divers medialandschap (1)
- Duurzamere campagne (4)
- Identiteit oproepen (5)
- Kostenreductie (7)
- Direct feedback in plaatje (9)
- Werkgelegenheid (juristen, prompter, etc.) (11)
- Nieuwe mogelijkheden -> actieve bijdrage (15)
- Creatief idee vertalen naar beeld (17)
- Specificeren op doelgroep, personalisatie (18)
- Verdwijnde banen (20)
- Flexibelere, bijsturen (21)
- Oneindige creativiteit (27)
- Controle, sturing (28)

Negatieve effecten

- Je moet zelf unbiased (2)
- Representatie nog ok? (3)
- Identiteit gebruiken zonder betrokkenheid (6)
- Ontstaan echoput -> teveel gebruik voor mooi mens (8)
- Ondermijnt creatief proces (10)
- Inbreuk op auteursrechten (12)
- Kijker weet niet wat echt is en wat nep (13)
- Meer kans op fouten (14)
- Afhankelijk van bedrijf, vendor lock-in (16)
- Verdwijnen banen (19)
- Achterblijvende bedrijven qua verdienmodel (22)
- Nadelig voor reclamebureau (23)
- Geen weerspiegeling meer van realiteit (24)
- Onmenselijk, fake mensen (25)
- Vraagt resources van producerend bedrijf (26)
- Verantwoordelijkheid verschuift naar andere partij (RTL) (29)
- Bescherming auteursrecht (30)
- Minder duurzaam (31)
- Versterkt risico's personalisatie (32)



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:

Waarden:

- **Diversiteit (representatie)** - 1, 2, 3, 8, 9, 32
- Digitale inclusie - 4
- **Waarheid** - 5, 6, 13
- Efficiëntie 7
- **Creativiteit** - 10, 15
- **Werkgelegenheid** - 11,19,20
- Eigendomsrecht - 12
- Betrouwbaarheid – 14, 16
- (zelf)Expressie - 17
- Privacy -12
- Onafhankelijkheid - 16
- Functionerende markt
- Effectiviteit (ook: personificatie: dat mensen zich persoonlijk - aangesproken voelen door een reclame) -18
- Flexibiliteit - 21
- Arbeidsvreugde- 26
- Continuïteit (van een organisatie) -22
- **Beheersbaarheid** – 25, 28
- Transparantie -

De vetgedrukte waarden werden door de groep als belangrijkste aangemerkt .



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:

- Technische beschrijving in model card of datasheet
- Vanrails -> bias detection tools
- Eigen model trainen, dat rekening houdt met Nederlandse situatie. Op basis van eigen materiaal en/of open source
- Gebruik van meerdere tools naast elkaar
- Vendor selectie
- Watermerk
- Integreren van origineel werk technisch makkelijker maken
- Reverse image search inbakken
- Proces van co-creatie technisch inbouwen
- Logs, audit trail
- Gebruiken system prompt
- Opletten op wat je erin stopt
- Diversere alternatieven
- Automatische screening
- Vierogenprincipe inbouwen in proces

Omgeving:

- Beleid formuleren door RTL over gebruik van generatieve AI
- Er is binnen RTL al een manifest, waarin bijvoorbeeld iets staat dat er geen herkenbare personen uit mogen komen. Dat moet steeds aangepast worden
- Er zou een gesprek moeten komen met de Reclame Code Commissie over dat manifest. Misschien kan het breder geldig worden, misschien moet het verder aangepast worden.
- Het invoeren van een quatum voor het gebruik van GAI-toepassingen
- Een diversiteitsscan bij het gebruik van GAI-toepassingen, bijvoorbeeld door het voor te leggen een breed samengesteld panel



- Organiseer ook een dialoog met het publiek, vertel dat je GAI hebt gebruikt en organiseer een (begeleidingsethiek) sessie voor een breder publiek
- Er zou transparant gemaakt kunnen worden wat je wel of niet ziet. Bijvoorbeeld: 'deze beelden zijn AI gegenereerd' of een merkje. Dit kan ook nog afhankelijk van het soort programma
- Er zou een doorlopend onderzoek moeten zijn naar de gevolgen van het gebruik van GAI op meerdere aspecten: duurzaamheid, diversiteit, werk etc.
- Er is omscholingsbeleid nodig voor mensen die hierdoor hun baan dreigen te verliezen
- De mensen die met GAI moeten werken moeten ondersteuning/training krijgen in het gebruik ervan. Uiteindelijk zal het ook onderdeel van de opleiding moeten worden, bijv. journalistiek.
- De ecologische voetafdruk van GAI zou gecompenseerd kunnen worden (een soort CO2 toeslaga o.i.d.)
- Brede toegang tot GAI-programma's als Midjourney, zodat veel mensen op deze wijze creatief kunnen worden
- Blijf ook 'klassieke' creativiteit stimuleren, anders verlies je veel.

Mens:

- Zorg voor AI-geletterdheid bij bevolking
- Bijscholing gebruikers – diversiteit, beheersbaarheid
- Omscholing
- Betrekken groepen die je gebruikt
- Human in the loop, in het hele proces
- Diverse groep mensen betrokken bij proces
- AI als middel voor de mens, niet als doel
- AI-visie in vacatures opnemen
- Nieuw personeel trainen
- Panels in blijven zetten voor evaluatie
- Bij gebruik van AI zorgen voor toegankelijke en beschikbare informatie



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 er in korte tijd veel wordt gerealiseerd. Daarnaast wordt de brede groep van deelnemers als meerwaarde ervaren. De combinatie van een duidelijke casus, enthousiaste deelnemers, en een aanpak die zorgt voor duidelijke uitkomsten in een korte tijd was dus een succes.



De deelnemers

1. **Daan Odijk** Hoofd Data & AI, RTL Nederland
2. **Marco Gijsen** Head of Creative Strategy, Ad Alliance
3. **Sander Heithuis** Communicatieadviseur, RTL Nederland
4. **Bianca van der Lee** Communicatieadviseur, BZK
5. **Sanne Eggengoor** Data scientist, RTL Nederland
6. **Juliska de Vries** Eventmanager, ECP
7. **Laurens van der Molen** Coördinator werkgroep Cultuur & Media, NL AIC
8. **Harm Steijn** Beleidsmedewerker, Gemeente Hilversum
9. **Iskaj Janssen** Data scientist, RTL Nederland
10. **Pieter van Koetsveld** Senior beleidsmedewerker, OC&W
11. **Saskia Blom** Manager Compliance & Legal, RTL Nederland
12. **Julia Bol** Team lead Corporate Communicatie en Merkstrategie, RTL Nederland
13. **Anne Thier** Senior beleidsmedewerker privacy
14. **Jasper Kars** Beleidsmedewerker Data & AI

Moderatoren:

Edwin Borst, Daniël Tijink, ECP, Platform voor de informatiesamenleving





Verslag workshop
Aanpak Begeleidingsethiek
Generatieve AI voor creatief
promotioneel materiaal

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

www.begeleidingsethiek.nl