

aanpak begeleidingsethiek

Verslag workshop

Aanpak Begeleidingsethiek

Schrijfhulp IRIS

Directie Dienst Publiek en Communicatie
Ministerie van Algemene Zaken

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

Initiatiefnemers: Eva van Kruchten, DPC, ministerie van Algemene Zaken
Moderatoren: Daniel Tijink, Aldert de Jongste (ECP)

Op donderdag 17 september 2025 organiseerde de Dienst Publiek en Communicatie (DPC) van het ministerie van Algemene Zaken op initiatief van Eva van Kruchten een sessie begeleidingsethiek over de casus 'Ondersteuning van AI-schrijfhulp IRIS bij de ontwikkeling van content voor rijkscommunicatie' Aan de workshop namen 18 deelnemers (zie bijlage voor specificatie) deel onder leiding van Daniel Tijink en Aldert de Jongste van ECP | Platform voor de Informatiesamenleving.

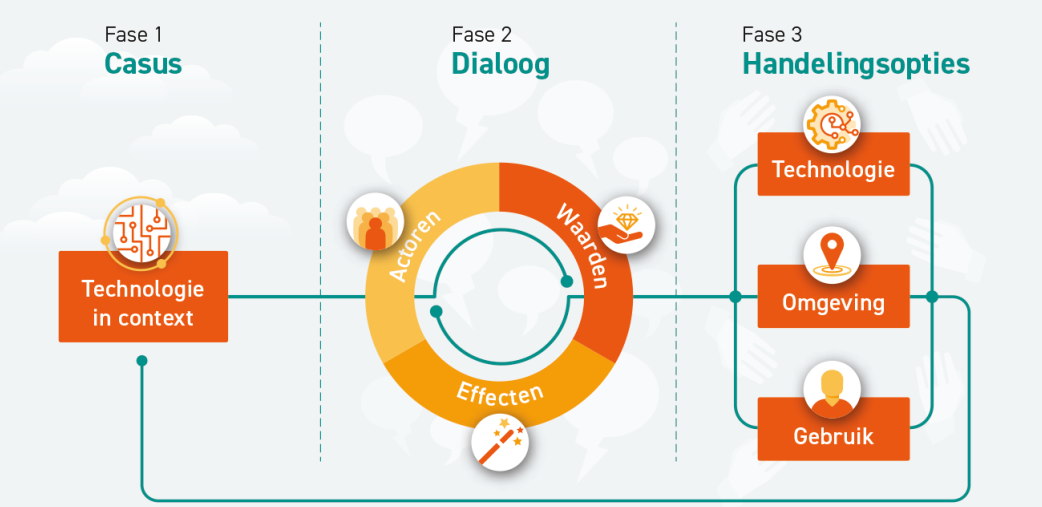
De doelstelling is tweeledig:

- 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 schrijfhulp IRIS.

Afgelopen maanden heeft een pilot gedraaid binnen het DPC waarbij redacteuren gebruik konden maken van een op maat gemaakte tool gebaseerd op een Large Language Model. De tool kreeg de naam IRIS mee en ondersteunde het schrijfwerk op verschillende manieren, zoals door teksten samen te vatten, suggesties te doen om een tekst te verbeteren, voorbeelden aan te dragen en te zoeken naar inhoud die al beschikbaar is op Rijksoverheid.nl over bepaalde onderwerpen. De output werd altijd door een redacteur gecontroleerd (human in the loop)

Fase 2

Dialoog

In deze tweede fase gaan de deelnemers in gesprek over wie er betrokken zijn bij de inzet en het gebruik van de schrijfhulp IRIS. 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 inzet van IRIS. De deelnemers aan tafel vertegenwoordigen al een deel van die actoren. Ze noemen de volgende betrokkenen:

- Redacteuren
- Ontwikkelaars
- Management
- Inhoudelijke verantwoordelijkheid
- Burgers/lezers
- Onderzoekers
- Het contactcentrum
- Compliance
- Beleidsmakers als gebruikers
- IT organisatie
- Niet rijksoverheden
- Inkoop
- Journalisten
- LLM-makers
- Hackers
- De Minister-President
- Taalexperts
- Redactiebeleid makers
- Data-analisten

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 schrijfhulp IRIS hebben

Positieve effecten

- Betere inrichting van de informatiehuishouding.
- Tijdswinst
- Eenduidige teksten, niet zo afhankelijk van wie schrijft
- We leren ook beter hoe AI in het algemeen werkt
- Je hebt met de schrijfhulp altijd een sparringpartner
- Het vervangt de zoekmachine
- Geeft goed overzicht
- Je krijgt nieuwe input voor je stukken, die worden rijker
- Grote datacollectie zijn makkelijker te doorzoeken
- We gaan mee met de ontwikkeling
- We worden/blijven een aantrekkelijke werkgever
- Het vak van redacteur gaat zich door ontwikkelen
- Het vak van redacteur wordt leuker
- Werken met supercoole tech is geweldig
- De kwaliteit wordt consistenter
- Minder redacteuren nodig

Negatieve effecten

- De webredacteuren worden luier en minder geconcentreerd, dat leidt toch tot minder goed werk
- Veel verspilling van grondstoffen, energie, etc
- De schrijfhulp is antropomorf, je denkt dat je met een mens te maken hebt ipv met een AI
- Tijdverlies, want tot een goede prompt komen en stukken aanpassen kost veel tijd
- Onjuiste informatie (hallucinaties)
- Onze betrouwbaarheid neemt af, komt toch uit de machine en dat betekent ook dat de impact afneemt
- De toon is heel gekunsteld, synthetisch
- Het gevaar dat we alles klakkeloos overnemen
- Reproduceerbaarheid is moeilijk, want het is een statistisch model
- Het werk van een redacteur wordt minder leuk
- De schrijfhulp zorg dat alles convergeert naar middelmatigheid
- Er zijn minder redacteuren nodig
- Invloed van big tech wordt (nog) groter
- Je krijgt minder kritische lezers.

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
Kwaliteit van dienstverlening
Efficiëntie
Duurzaamheid
Plezier
Zelfontwikkeling
Menselijkheid
Duidelijkheid
Innovatie/wendbaarheid
Betrouwbaarheid/Zorgvuldigheid
Geloofwaardigheid
Autonomie
Neutraliteit/objectiviteit
Veiligheid
Inclusiviteit

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. De getallen tussenhaakjes zijn onderdeel van de prioritering die na de handelingsopties is aangebracht en kort beschreven is onder het volgende kopje.

Technologie:

Kwaliteit

- Maak copy-paste onmogelijk, zo dwing je redacteuren (ondersteund) eigen tekst te typen
- Zorg voor serieuze kalibratie: welke input leidt tot welke output, laat mensen daarover oordelen
- Aan de achterkant van een tool zijn allerlei schuifjes, zorg dat die bewust ingesteld worden (1)
- Maak een bewuste keuze voor de te gebruiken LLM op basis van relevante waarden.
-

Efficiëntie

- Geef mensen een maximaal aantal queries per dag
- Zorg voor een soepele manier om in te loggen
- De interface van de schrijfhulp moet intuïtiever worden voor gebruikers (1)
- De schrijfhulp past zich aan aan de behoeften van de gebruiker (3)
- Maak duidelijk wat de schrijfhulp nu precies oplevert

Betrouwbaarheid – zorgvuldigheid

- Zorg voor een monitoringsfunctie, die bij de bron checkt of de informatie wel klopt (5)

- Laat de schrijfhulp aangeven wanneer hij ergens onzeker over is of het niet weet (6)
- Bronverwijzingen zijn direct aanklikbaar (6)
- Sluit onbetrouwbare bronnen uit (1)

Menselijkheid

- Geef de schrijfhulp geen menselijke naam, je werkt immers juist niet met een mens. Misschien gewoon ‘schrijfhulp’ (of LLM-SH) (5)
- Laat de schrijfhulp een schrijfhulp zijn, dus de redacteuren helpen beter te schrijven en niet zomaar teksten genereren
- Zorg dat de resultaten uit een schrijfhulp niet te gelikt zijn, je ze blijven wantrouwen (1)

Omgeving:

Kwaliteit

- 4 ogen principe + inhoudelijke check bij de ministeries (6)
- X% blijft door mensen geschreven (3)
- Evalueren van de kwaliteit van de output, mede aan de hand van richtlijnen die we nog moeten formuleren

Efficiëntie

- Instructies voor prompt optimalisatie, periodiek ook analyse en brede implementatie/vernieuwing prompts (2)
- Periodiek analyseren hoeveel energie het systeem gebruikt en hierover rapporteren
- Trainingen aanbieden, welke functies gebruik je waarvoor (9)
- Dynamische prompt bibliotheek, creëer een feedback loop en deel best practises

Betrouwbaarheid – zorgvuldigheid

- Laat een bot periodiek de betrouwbaarheid checken (2)
- Creëer de mogelijkheid tot het aangeven als Iris ongelijk heeft (2)

Menselijkheid

- Creëer een AI- label zodat duidelijk is voor lezer hoe iets gecreëerd is (1)
- Onderzoek de resultaten met behulp van echte mensen en synthetische persona

Mens:

Betrouwbaarheid

- Zorgen dat gebruikers de bot kritisch blijven bevragen (6)
- Gebruikers instrueren om bronnen te blijven checken (4)
- Altijd transparant zijn over welke bronnen er gebruikt zijn en wat er mee gedaan is (1)
- Stakeholders goed meenemen en laten weten wat je als gebruiker nodig hebt

Kwaliteit

- Mens/gebruiker heeft altijd het laatste woord (2)
- Alle betrokkenen goed meenemen in hoe de tool werkt en hoe je deze moet gebruiken (5)
- Blijf de tool doorlopend toetsen en doorontwikkelen. Dit moet ook door mensen gebeuren.

Efficiëntie

- De redacteur is beslisser over waar wel of geen AI gebruik. Dit om onnodig AI gebruik tegen te gaan (2)
- Prompts en hulpbronnen opslaan en beschikbaar maken (3)
- Goed handleidingen maken en bijhouden
- Borgen dat tijdswinst door AI gebruikt wordt voor kwaliteitsverbetering

Menselijkheid

- Blijf ook na introductie testen met echte mensen, op teksten, gebruik en processen (1)
- Mensen/gebruikers trainen op goed inzet (2)
- Ontwerp en doorontwikkeling door mensen, redacteuren en lezers
- Voorkom aannames door het continue toetsen van teksten

Prioritering

Nadat de handelingsopties uit de subgroepjes plenair aan de collega's zijn gepresenteerd krijgen alle deelnemers de gelegenheid om de voor hen belangrijkste handelingsopties te bepalen. Alle deelnemers kunnen twee handelingsopties uit kiezen per domein (technologie, omgeving en mens). In bovenstaande overzicht van handelingsopties zijn de geturfde voorkeuren te zien als cijfer tussen haakjes. Het was een vluchtige exercitie. Met de diversiteit aan stakeholders aan tafel is een flinke spreiding van waar men de prioriteit legt vanzelfsprekend.
Handelingsopties die de waarde betrouwbaarheid vergroten zijn veel gekozen als het gaat om aanpassingen van de Technologie. De handelingsopties gaan over aanpassingen aan de toepassing waardoor de juistheid van de uitkomst beter gegarandeerd wordt of beter controleerbaar is. Bijvoorbeeld door de optie: "Laat de schrijfhulp aangeven wanneer hij ergens onzeker over is of het niet weet"
In handelingsopties voor de Omgeving speelt waarde kwaliteit maar ook efficiëntie een belangrijke rol. Die laatste is vooral ingegeven door de handelingsoptie "Trainingen aanbieden, welke functies gebruik je

waarvoor”, deze springt eruit qua populariteit, wat mogelijk verklaarbaar is door het zeer generieke karakter waardoor deze optie potentieel invloed heeft op alle waarden. Binnen het domein Mens is ook de meeste prioriteit gegeven aan handelingsopties die de waarde betrouwbaarheid raken. De twee meest gekozen opties gaan over het instrueren van gebruikers de uitkomsten vooral kritisch te blijven benaderen.

Terugblik en afronding

Aan het einde van de workshop wordt een rondje gedaan langs alle deelnemers met grofweg twee vragen, wat vond je van de sessie en wat vond je van de opbrengst inhoudelijk. De volgende zaken worden genoemd.

- Mooi hoe informatie uit de groep getrokken wordt tijdens de workshop
- Mooi om wat we in de wandelgangen bespreken nu in een groep bij elkaar te bespreken
- We zitten wel op een lijn, erg benieuwd wat we nu gaan doen
- Interessant alle verschillende perspectieven bij elkaar
- Leuk te zien welke waarden we pakten, goed om deze verder ook mee te nemen tijdens ons werk
- Moeten mee met vaart der volkeren, nu hebben we het goed over uitgangspunten
- Goed duidelijk dat we moeten nadenken ook over wat we naar buiten brengen
- We moeten ook met de andere waarden aan de slag, mogelijk ook met andere stakeholders
- Hoe gaan we dit nu allemaal concreet maken
- Vanuit design thinking met de groep in de pilot iets opgeleverd, nu is alles weer open, iedereen heeft de behoefte om weer te vernauwen en zaken concreet te maken

De deelnemers

De deelnemers bestonden uit mensen met de volgende rollen binnen de organisatie;

1. Redacteur
2. Onderzoeksadviseur
3. Adviseur inclusie
4. Kennisadviseur
5. IB & Privacy adviseur
6. Programmeur
7. Architect
8. Medewerker contactcenter
9. Beeldadviseur

Moderatoren:
Daniel Tijink en Aldert de Jongste (ECP)

aanpak begeleidings ethiek

ECP Platform voor de InformatieSamenleving

Verslag workshop
Aanpak Begeleidingsethiek
Naam workshop

Meer informatie over de Aanpak Begeleidingsethiek, waaronder dit verslag, vindt u op **www.begeleidingsethiek.nl**

Foto omslag: Shutterstock.com