

aanpak begeleidingsethiek

Verslag workshop
Aanpak Begeleidingsethiek

Anonimiteit op social media

Meta en Ministerie van BZK

ECP Platform voor de InformatieSamenleving

Ministerie van Binnenlandse Zaken en Koninkrijksrelaties

Meta

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: Dorothy Adobea, Claire Gersen (BZK)
Moderatoren: Daniël Tijink (ECP), Marieke Peeters (Mooncake AI, HU)

Op 27 november 2023 organiseerde het ministerie van BZK en Meta op initiatief van Dorothy Adobea en Claire Gersen een sessie begeleidingsethiek over de casus 'Anonimiteit op Social Media'. Aan de workshop namen 15 deelnemers (zie bijlage voor specificatie) deel onder leiding van Daniel Tijink (ECP) en Marieke Peeters (Mooncake AI, HU).

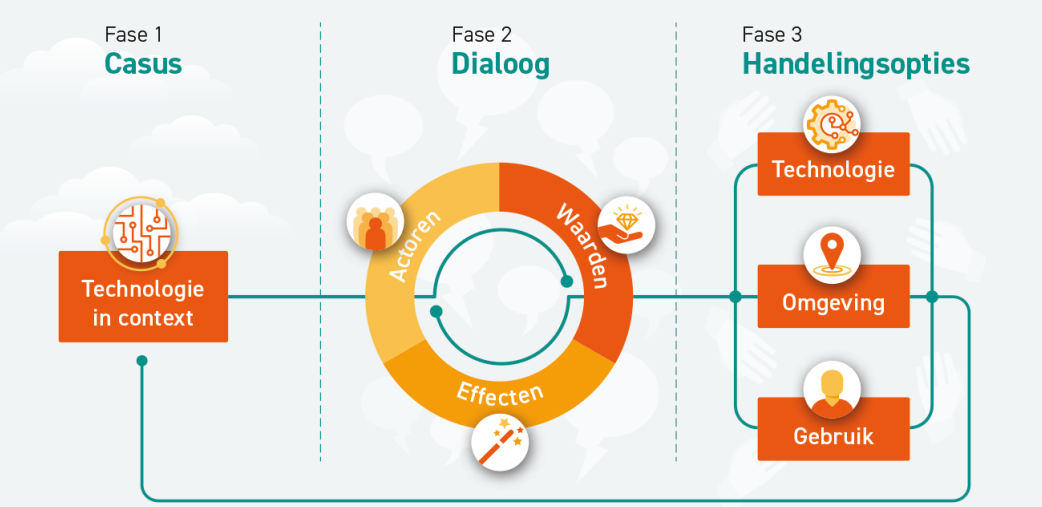
De doelstelling is tweeledig:

- 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

Dorothy introduceerde het vraagstuk vanuit BZK: “Wat is ons waardenoordeel over anonimiteit online?” Daar wordt ook onderzoek naar gedaan door onderzoekers van de universiteit Wageningen. We bevinden ons nu bij een van de vier sessies met de aanpak begeleidingsethiek (ABE). Die hebben het gebruik van verschillende technologieën, instrumenten, die invloed hebben op anonimiteit en zichtbaarheid als onderwerp. Dit is de eerste sessie van een cluster aan sessies – hierna vinden er nog 3 andere sessies ABE plaats die ook te maken hebben met online anonimiteit. Vandaag ging het vooral over de authentic name policy van Meta, wat betekent dat mensen geacht worden hun echte naam te gebruiken op Facebook.

Vervolgens nam Edo Haveman (Meta) het stokje over, die de case toelichtte. Hij vertelde meer over de Authentic Name Policy (ANP) bij Meta, met name op Facebook. Deze policy volgt uit het feit dat Facebook letterlijk een smoelenboek is, maar heeft als waardevolle bijvangst dat het de veiligheid op het platform kan vergroten. Hij gaf aan dat het geen silver bullet is, maar wel een behulpzaam element binnen een bredere aanpak. Veiligheid op het Facebook platform wordt ondersteund door een aantal pilaren:

- Beleid, regels, en richtlijnen, zoals de Richtlijnen voor de Community waarin alles beschreven is wat wel en niet toegestaan is op Facebook te delen, waaronder de ANP[1].
- Handhaving van het beleid door moderatoren, in combinatie met AI-oplossingen om het platform te monitoren, en overtredende content te signaleren en eventueel (automatisch) te verwijderen.
- Verschillende tools, waaronder de mogelijkheid voor gebruikers om content als problematisch of overtredend aan te merken, via formulieren om hulp te vragen, of te escaleren.
- Partnerships met lokale organisaties zoals OffLimits of meldpunt online discriminatie om te leren en de veiligheid op het platform aan te scherpen.
- Feedback, van mensen op het platform maar ook door bijvoorbeeld in gesprek te gaan met experts en/of beleidsmakers (zoals BZK, maar ook de onderzoekers van de WUR die vandaag aanwezig zijn).

ANP beschrijft dat je geacht wordt op het platform de naam te gebruiken waaronder je bekend staat in het dagelijkse leven. Het hoeft niet de naam op je paspoort te zijn, maar wel de naam waaronder je bekend bent bij familie en/of vrienden. Er zijn regels die impersonatie (je voordoen als iemand anders) tegen gaan en er is een eis om authenticiteit. Ook gaan ze het hebben van meerdere accounts tegen (op Instagram is dit overigens anders en kun je wel een account voor bijvoorbeeld je kat hebben). Het moet een gangbare en logische naam zijn. Schuttingtaal of scheldwoorden in je naam zijn niet toegestaan. Ook regelmatige

[1] https://transparency.fb.com/policies/community-standards/account-integrity-and-authentic-identity/

naamwisselingen, of het gebruik van een naam die al door iemand anders in bezit is, zijn een reden om iemand nader te onderzoeken. Je mag je niet voordoen als iemand anders of proberen anderen te misleiden of op te lichten. Het hebben van meerdere accounts, bijvoorbeeld om te spammen, wordt ook actief tegengegaan. Wanneer er een vermoeden is van overtreding van de regels, kan afhankelijk van de overtreding gevraagd worden om een videoselfie of identificatie. In een aantal gevallen wordt mensen gevraagd zichzelf te identificeren middels een overheids-ID, bijvoorbeeld wanneer je een abonnement wilt op Meta Verified (o.m. blauw vinkje), of wanneer je politieke advertenties of advertenties over maatschappelijke onderwerpen wilt plaatsen.

Fase 2
Dialoog

In deze tweede fase gaan de deelnemers in gesprek over wie er betrokken zijn bij de inzet en het gebruik van authentic name policy. 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:

- Burgers
    - Ouders van minderjarigen
    - Publieke / bekende mensen
    - Wetenschappers
    - Journalisten
- Gebruikers
    - Kinderen
    - Publieke / bekende mensen
    - Wetenschappers
    - Journalisten
    - Communicatieprofessionals
    - Activisten
    - Influencers
- Bedreigde mensen
    - Slachtoffers van geweldsituaties
    - Slachtoffers van stalkers
- Organisaties
    - Belangenvertegenwoordigers
    - Privacywaakhonden
    - Bits of freedom
    - Senioren organisaties
    - AP / IDPC / EDPD
    - ACM
    - Trusted flaggers
        - Meldpunt online discriminatie
- Overheden als gebruikers
    - Gemeente
    - Ministerie BZK
    - Ministerie J&V
- Politie / OM (handhaving)
- Bijzondere opsporingsdiensten

    - FIOD
    - AIVD
    - MIVD
    - NCTV / NCSC
- Meta
    - Instagram
    - Facebook
    - Moderatoren
    - Directie / management
    - Oversight board
- Wet en beleid
    - EU (VSA, GDPR)
    - Nationaal
    - Gemeente
- Commerciële partijen / adverteerders
- Buitenlandse mogendheden

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 het gebruik van Real Name Policy hebben

Positieve effecten

1. Keuze in naam (hoeft niet zelfde te zijn als paspoort) - vrijheid
2. vertrouwen in Meta
3. Onderling vertrouwen tussen gebruikers
9. Meta weet veel – controlerende functie
10. Afzender / bron bekend (attribueerbaar – weerbaarder)
11. hogere drempel haatmail
12. minder spam en daardoor ook minder CO2 verbruik (duurzaamheid)
14. geen langdurige kwetsbaarheid
16. Geen 10 miljard paspoorten in database meta
17. Ontlasting van toezicht / politie
19. Echte connectie (met bijvoorbeeld oude bekenden)
21. Identiteitsvorming, bijv. trans of adoptie
23. Dichtbij natuurlijk persoon (recht)

Negatieve effecten

4. Kwetsbaar voor persoonlijke aanvallen
5. Minder bereid om deel te nemen aan discussies (inclusiviteit)
6. Traceerbaarheid kan beter
7. Drempelverhogend om toe te treden
8. Datamisbruik rond naam
13. Systeem werkt mogelijk niet altijd goed
15. Transparantie over opvolging / traceerbaarheid
18. Afstand tot toezicht
20. Relatie met de echte wereld makkelijk (pesten enzo)
22. Schijnveiligheid – alertheid op fake accounts

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:

Veiligheid / zelfexpressie / stem	1, 4, 5, 8, 21
Vertrouwen / betrouwbaarheid	2, 3, 6, 11, 14, 22
Inclusiviteit	5, 7, 21
Toegankelijkheid	7
Veiligheid	6, 9, 10, 11, 13, 14, 15, 16, 17, 18, 21, 22, 23
Privacy	6, 8, 9, 10, 11, 15, 16, 17
Toerekenbaarheid	8, 9, 10, 6, 11, 15, 16, 17, 18, 20, 21, 23
Duurzaamheid	12
Transparantie	15, 16
Bescherming (kwetsbaarheid)	8, 16, 4, 10, 11, 17, 18, 20, 21, 22, 23
Connectie	19
Verantwoordelijkheid	22, 9, 17, 18, 6, 8, 13, 15, 22
Zelfbeschikking / autonomie	
Waarachtigheid	

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:

- Luchthavenmodel: een deel waar iedereen binnen kan en daarna een deel PPS, waar andere toegangscodes en afspraken gelden. Waar overheid en platformen samenwerken aan een inlog met DigiD, waar je vervolgens onder eigen naam kunt surfen, wel herleidbaar, maar potentieel anoniem
- Anonieme discussiezone: binnen platform bepaalde plekken, waar gemarginaliseerde groepen toch anoniem kunnen praten
- Bevestiging door derden (connecties verifiëren je identiteit)
- Vindbaarheid via cirkelconnecties (linked-in model) – profiellink delen wanneer je geen connectie bent, als je in grotere discussies meedoet zien mensen die geen connectie zijn jou als anoniem
- Dialoog AI/Algoritme die/dat commentaar geeft op beleefdheid of onaardige berichten en die waardevolle bijdragen hoger in de comments plaatst (Reddit)
- Unique name policy (identifier) – sociale identificatie
- Web scraping – real name policy – niet iedereen zomaar webscraping, moeilijker maken vanuit platform, bijv. onderzoekers wel toegang geven, maar anderen niet.

Omgeving:

- Probleem dat hier ligt – soms naïef in fraude trappen. Beleid dat kijkt naar toegankelijkheid en veiligheid van verplichtstelling. Vergelijkbaar met cookie waarschuwingen.
- Mediawijsheid vast onderdeel van het curriculum – wat geef je weg wanneer je onder je eigen naam een like geeft aan je favoriete politicus
- Keuze voor anonieme account -> transparantie over dat een account niet geverifieerd is en het niet duidelijk is wie er achter zit.
- Onderzoek doen – wat is het effect van al dan niet authentic name policy? Gedragen mensen zich echt anders wanneer ze hun eigen naam gebruiken?
- Ook onderzoek naar best practices om anonimiteit te waarborgen.

Mens:

- Informatie en toelichting op het waarom van ANP vanuit het idee dat het helpt voor acceptatie en bewustwording als je weet waarom, vanuit welke waarden hiervoor gekozen is,
- Leerdoelen over anonimiteit/zichtbaarheid meenemen via onderwijs / mediawijsheid / positief gedrag online – hoe ga je met elkaar om? Wat zijn de normen?
- Nudges vanuit Facebook – Dus kleine berichtjes bij 'agressieve' berichten: bijvoorbeeld: 'dit kan aardiger'. Suggesties voor handelingsmogelijkheden. Hoe kun je een 'betere gebruiker worden'
- Ouders opleiden in hoe ze hun kinderen kunnen coachen
- Feedback richting facebook via gebruikerspanel
- Fatsoen digitaal en niet-digitaal. Die twee beïnvloeden elkaar, dus ook aandacht voor hoe je met elkaar omgaat e wa het doet in het echte en in het digitale leven als je onprettig wordt behandeld.
- Bewustwording/begeleiding van administratie/moderatie. De mensen binnen facebook die het proces van real-name policy faciliteren – let er op alles wat je noteert doet er toe, is van belang, heeft impact.
- Goed voorbeeld geven – bijvoorbeeld door de overheid, influencers "ik ben mezelf online – jij toch ook?"

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.

Zo gaven deelnemers aan dat het fijn was dat er toch ook nieuwe inzichten waren ondanks dat het een zombie onderwerp is dat elke zoveel jaar weer de kop opsteekt. Mensen waren ook benieuwd naar het vervolg, en hadden het idee dat de geïdentificeerde waarden ook wel breder toepasbaar zijn in het digitale domein. Ook vond men het fijn dat je als mensen onder elkaar het heel goed over dit onderwerp kunt hebben in deze setting, door het vanuit verschillende posities te bespreken. Bepaalde waarden bleken belangrijker dan andere waarden, er was geen hard tegen hard discussie. De effecten die genoemd werden waren zowel positief als negatief – goed om dat door te hebben. Vanuit de rol van initiatiefnemer gaf Dorothy aan heel blij te zijn met de sessieomdat technologie en ethiek zo naadloos samenkwamen: “dit is goud” – we kunnen dit ook verder implementeren en meenemen in besluiten die we nemen. Anonimiteit komt veel voorbij in discussies over opsporing, blijft groot punt. Anonimiteit is voor opsporing nooit volledig, wij kunnen bij veel gegevens komen. Bij online pesten is wel vaak de vraag of we daar niet meer mee kunnen of mogen.

De deelnemers

1. **Axel van Pel** MOD
2. **Dorothy Adobea** Ministerie BZK
3. **Daan Hofland** Publyon
4. **Emma Turkenburg** WUR
5. **Dorijn Boogaard** Jong ECP
6. **Sanne Kruikemeier** WUR
7. **Christian Verhagen** VKA
8. **Tessa van Hemst** Publyon
9. **Edo Haveman** Meta
10. **Eric Brendel** ANBO
11. **Claire Gersen** Ministerie BZK
12. **Gerbrand Mijzen** Politie
13. **Gijs van Beek** Textgain
14. **Yoni Dekker** Ministerie J&V

Moderatoren:
Daniel Tijink (ECP), Marieke Peeters (Mooncake AI / HU)

aanpak begeleidings ethiek

ECP Platform voor de InformatieSamenleving

Verslag workshop
Aanpak Begeleidingsethiek

Anonimiteit op social media

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

Foto omslag: deviantart.com (https://www.deviantart.com/004946red/art/Mosaic-identity-2-305020375)