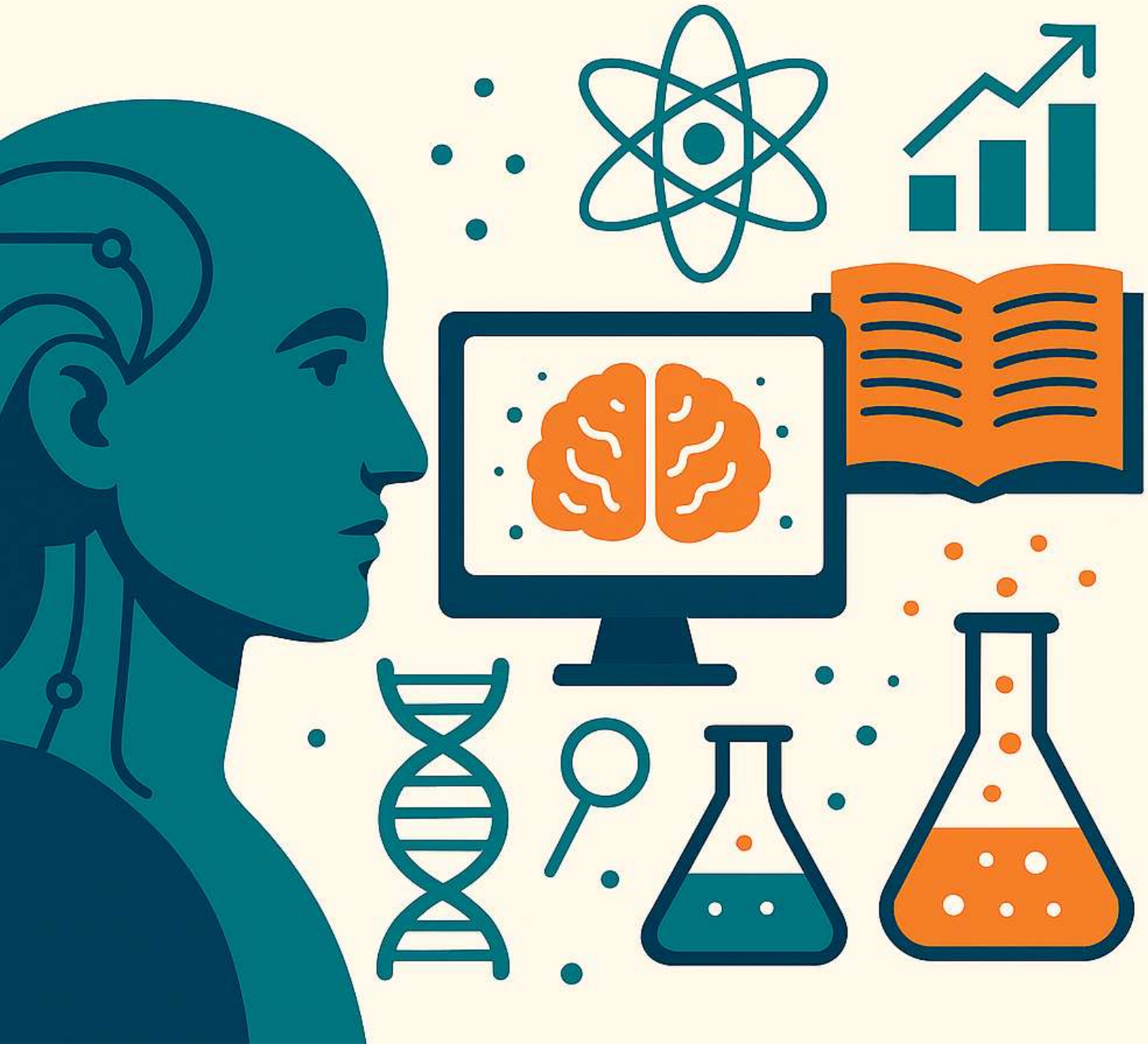

aanpak begeleidings ethiek

Verslag workshop

Aanpak Begeleidingsethiek

Wetenschappelijke kennisverspreiding in het generatieve AI-tijdperk.

Over Informatie Gesproken

SURF

ECP Platform voor de InformatieSamenleving

Voorwoord

De toenemende digitalisering en zeker ook AI brengen 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 gaan verschillende stakeholders als gebruikers, ontwikkelaars, beleidsvormers met elkaar in gesprek. Er is gezamenlijk gekeken welke waarden we in het digitale domein belangrijk vinden en hoe er vanuit die waarden kan worden gehandeld. Daarbij is gekeken naar handelingsopties vanuit de techniek, de collectieve afspraken en het individueel gedrag. Er worden tijdens de sessie verschillende handelingsopties gegenereerd. Er is hierbij niet gezocht naar consensus, maar juist naar het komen van veelsoortige mogelijkheden.

Workshop aanpak begeleidingsethiek

Initiatiefnemers: Irene Nooren (SURF)
Moderatoren: Stephan de Valk en Daniël Tijink (beiden ECP)

Op 28 oktober 2025 organiseerde SURF, in samenwerking met de Maatschappelijke Coalitie Over Informatie Gesproken een sessie begeleidingsethiek over wetenschappelijke kennisverspreiding in het generatieve AI-tijdperk. Aan de workshop namen 18 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 van elkaar, vergoten van onderlinge begrip
- Leren over de Aanpak Begeleidingsethiek
  - Hoe kan een organisatie deze methodiek gebruiken rondom ethische kwesties en trajecten van in gebruik neming van technologische toepassingen?

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

Wetenschappelijke kennisverspreiding in het generatieve AI-tijdperk

De introductie van generatieve AI en Large Language Models (LLM's) in de wetenschap dwingt de wetenschappelijke gemeenschap te reflecteren op fundamentele invloeden zoals kennisgezag en kwaliteit van wetenschappelijke output. In deze casus nemen we het effect van generatieve AI op het wetenschappelijke publicatieproces onder de loep, specifiek in de validatie en disseminatie van wetenschappelijke output. Deze output vormt de basis voor verder onderzoek, beleidsvorming en praktijktoepassing. Hoe blijft de rol van de wetenschap in het waarborgen van kwaliteit en archivering van kennis geborgd?

Generatieve AI heeft het potentieel om elk aspect van het wetenschappelijke publicatieproces te veranderen. Zodra een onderzoeker een manuscript bij een uitgever indient, wordt een manuscript door een *editor* beoordeeld op kwaliteit. AI kan en wordt hier al ingezet voor een initiële screening en zelfs het *peer review* proces zelf. De *peer review* is traditioneel de taak van redactiegroepen. Bij groeiende wetenschappelijke output is generatieve AI een technologische oplossing voor de analyse en verwerking van data wanneer dit de menselijke capaciteit overstijgt.

Bij acceptatie wordt het artikel op basis van open science principes *open access* gepubliceerd voor optimaal hergebruik. Teksten kunnen door generatieve AI toegankelijker gemaakt worden voor burgers en burgerwetenschappers door het maken van ingekorte varianten, waarbij kwaliteit door de groep ter discussie wordt gesteld. In het toegankelijk maken van wetenschappelijke output voor het brede publiek hebben journalisten en wetenschapscommunicatoren traditioneel een rol.

Citatie is de levensader van het wetenschappelijke proces. Bij het hergebruik van kennis, vaak vrijgegeven onder een *Creative Commons* (CC) BY licentie, is naamsvermelding van de

auteur verplicht. Generatieve AI kan niet automatisch de oorsprong van informatie traceren om aan de academische eis van correcte bronvermelding te voldoen. Daar is het technisch niet voor ontworpen. Het resulteert in verlies van waardering en zichtbaarheid voor de wetenschappelijke bijdragen.

Als onderdeel van de case werd het landschap geschetst waarin generatieve AI, en specifiek LLM's zich ontwikkelen. Wetenschappelijk output heeft een hoge kwaliteit en is daarom gewild in het gebruik van trainingsdata. Daar ligt een technische en juridische uitdaging. Het is technisch onmogelijk om het web-scrapen van content door grootschalige AI bots tegen te houden. Het niet-ethisch scrapen van academische artikelen met bijvoorbeeld een CC non-commercial licentie is daarin ongewenst. De beperkte juridische grond voor generatieve AI is omstreden door de onduidelijkheid hoe grootschalige tekst- en datamining (waar genAI onder valt) zich verhoudt met auteursrecht.

De rol van de wetenschap is het leveren van gevalideerde en verifieerbare input aan de maatschappij en beleidsvorming, waarbij we traditioneel vertrouwen op menselijke kennis. Generatieve AI voegt een nieuwe 'denk laag' toe aan het wetenschappelijke proces. Hierbij wijst de groep op het antropomorfiseren van de handelingen van generatieve AI. Wat zijn de effecten van deze technologie en hoe zijn de rollen van mens en machine te herdefiniëren om de integriteit en het vertrouwen in wetenschappelijke output te blijven garanderen?

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 generatieve AI. 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:

- Bibliotheken
- Journalisten
- Wetenschappers
- Wetenschappelijke uitgevers
- Universiteiten
- Kennisinstellingen
- Overheden
- Burgers
- Beleidsmedewerkers
- Industrie
- Studenten
- Musea
- Actiegroepen
- Patiëntenverenigingen
- SURF
- Redactiegroepen
- Peer-reviewers
- Editors
- Auteurs
- Politieke actoren

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 bij kennisverspreiding van wetenschappelijke kennis hebben. De groep merkt op dat sommige effecten zowel positieve als negatieve kanten hebben. Voor het overzicht is hieronder wel een indeling gemaakt tussen de twee contrasten.

Positieve effecten

- Nog meer verspreiding van kennis
- Beter de achterliggende bedoeling van de gebruiker in beeld krijgen
- AI is heel flexibel, je kunt makkelijk het juiste format voor verspreiding van onderzoek kennis krijgen
- Meer tijd voor andere zaken
- Grotere productiviteit
- Makkelijker carrière maken, als je AI goed begrijpt en kunt inzetten
- Goedkoper, het kost minder onderzoekers, veel data is sneller beschikbaar etc.
- Ook nu bestaat er allerlei bias, door die te vergelijken door AI krijgen we daar meer inzicht in
- Toegankelijkheid van wetenschappelijke kennis wordt vergroot
- Wetenschappelijke kennis wordt toegankelijker voor hen die geen Engels spreken (China, Global South)
- Het draagt bij aan het oplossen van de reviewer crisis
- AI kan extra creativiteit inbrengen
- De kwaliteit van verspreiding van onderzoek kennis gaat omhoog
- Er zijn snellere en betere transparantie checks mogelijk
- Er is een betere detectie mogelijk of onderzoek/een vinding nieuw is
- Er is patroonherkenning mogelijk op veel grotere schaal
- Meer transdisciplinariteit
- Reviewen gaat sneller en beter

Negatieve effecten

- Minder kwaliteit
- Klimaatimpact
- Meer fraude mogelijkheden
- Afhankelijkheid van big tech wordt (nog) groter
- Onzekere betrouwbaarheid (hallucinatie, ‘state of the art data’ niet meegenomen in training, link met de bron verdwenen)
- Onduidelijk waar de verantwoordelijkheid ligt
- De waardering van onderzoekers wordt minder
- Er ontstaat een angst om open te delen, omdat AI dat allemaal ‘inpikt’ voor trainingsdoeleinden
- Er komt nog meer kennis en er is al zo veel. We moeten naar beter minder in plaats van meer-meer.
- Devaluatie van wetenschappelijke teksten als kennis
- Taal en kennis worden vlakker, AI gaat voor het gemiddelde, niet voor het beste of mooiste
- Menselijke creativiteit neemt af
- Poisening the well (de bron van kennis wordt door misinformatie vertroebeld en vervuilt zo het bestaande wetenschappelijke kennissysteem)
- Fundamentele vernieuwing wordt onmogelijk
- Er wordt minder zorgvuldig geformuleerd

- Er komen betere samenvattingen
- Er komen nieuwe verdienmodellen

- Er komt een 3e verdienmodel bij. Naast de wetenschap en de uitgevers
- Meer gebruik van datacenters
- Bestaande bias wordt versterkt (bv Noord-Zuid)
- Oncontroleerbaarheid neemt toe
- Het wordt minder transparant, black box
- Grote twijfel of de burger/iedereen nog wel mee kan doen
- Trainingsdata lopen steeds meer achter
- Antropomorfisering, we denken dat AI menselijk is

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		
Kennisverspreiding	Efficiëntie	Creativiteit
Kwaliteit van onderzoek	Werkplezier	Vakmanschap
Duurzaamheid	Verantwoordelijkheid	Verdienvermogen
Flexibiliteit	Erkenning	Menselijkheid
Gebruiksgemak	Transparantie	Waarheid
Rechtvaardigheid	Toegankelijkheid	Wetenschappelijke integriteit
Reproduceerbaarheid	**Betrouwbaarheid**	**Onafhankelijkheid**

Ook werd de deelnemers gevraagd welke waarden zij het meest relevant vinden. Daarover ontstond een goede dialoog waarbij er drie waarden zijn gekozen: kennisverspreiding, onafhankelijkheid en betrouwbaarheid. Deze waarden zijn leidend geweest bij het komen tot handelingsopties.

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

Onafhankelijkheid (behoud van wetenschappelijke autonomie)

- Als risico mitigerende maatregel: richt een extra rol in (mens of machine) die een objectieve check kan doen op de inhoud. Per case of domein is te bepalen wanneer dit ingebouwd moet worden en wie (of wat) deze rol heeft. Dit kan een domeinexpert zijn of een AI bot. Een AI bot kan bijvoorbeeld prima ingezet worden voor het valideren van standaard kennis.
- Stel vast wie (menselijke rol of instantie) bepaalt wat belangrijk is, bijvoorbeeld hoe een genAI applicatie zoekresultaten prioriteert op basis van trainingsdata. Wie bepaalt de normen?
- Breng 'humans in de loop' waar het nodig is om regie te houden op output.
- GenAI modellen en applicaties moeten open toegankelijk en voor iedereen te gebruiken zijn.
- Acteren tegen schaarste die de Big Tech creëert: Big Tech heeft nu controle over hoe de technologie er uit moet zien, heeft de middelen en daarmee de macht. Hier moet interventie plaatsvinden.

Betrouwbaarheid

- Containerisatie (het ‘verpakken' van de software, data, model en pipeline/workflow inclusiefprompts in een ‘container’) voor reproduceerbaarheid van het kennisproductie proces.
- Gecontroleerde en vertrouwde omgeving (bijv. SURF) waarin containers beschikbaar worden gemaakt.
- Het genAI model en de gebruikte data moet open en transparant zijn, bewijsbaar kwalitatief ondermaats of juist goed; geen black box.

Kennisverspreiding

- Digitaal watermerk waarmee je kunt zien dat content AI gegenereerd is.
- Bronmateriaal citeren en attributie van (samenvattingen van) wetenschappelijke output ter erkenning van de auteur/onderzoeker.
- Creëer vertaling AI tools die academische taal in lekentaal en praktijkinzichten kan omzetten.
- Multimodale AI applicaties en modellen die diverse type data zoals publicaties en video's faciliteren.

Omgeving/Beleid

Onafhankelijkheid

- 'Terugscrapen' door GPT-NL, de makers van GTP-NL gaan bij de Big Tech langs om hun data 'af te pakken'.
- EU-versie/Open source versie van genAI modellen i.p.v. Big Tech (publiek gefinancierd).
- Samenwerking universiteiten in de EU.
- De opslag van data, van de LLM, de researchdata en de publicaties zou in Europa moeten plaatsvinden.

Betrouwbaarheid

- In de opleiding kritische AI-vaardigheden opnemen. Weten welke effecten bepaalt gebruik heeft, en welke toepassingen ethisch beter zijn dan anderen.
- Eisen aan transparantie van gebruik genAI.
- Kwalitatieve erkenning en waardering: Nu gaat het vaak om de hoeveelheid artikelen en publicaties, maar AI kan beter helpen de kwaliteit en uniciteit in te schatten. Dat betekent ook dat de prestatiedruk naar beneden kan en de kwaliteit omhoog.
- Beleid op technische (genAI-)tools voor check op betrouwbaarheid
- Collectieve peer review
- ('Draconisch'): We verbieden het gebruik van AI voor verspreiding en waar dat toch gebeurt, vindt uitstoten uit de wetenschappelijke community plaats.

Kennisverspreiding

- Bij elk (wetenschappelijk) artikel is de wetenschapper verantwoordelijk voor minimaal twee andere verspreidingsvormen dan het wetenschappelijk artikel. Zo blijft de wetenschapper zelf in controle over de inhoud én wordt deze gedwongen de kennis voor een breder publiek toegankelijk te maken.
- Er wordt veel kennis verspreid die niet klopt, maak een tool om effectieve kennisverspreiding te kunnen toetsen.

Mens

Onafhankelijkheid

- Gebruik alleen ethische LLM's, bv uit Europa.
- Als we het hebben over AI-vaardigheden (*literacy*) dan hoort daar ook bij bewustzijn van de onafhankelijkheid van AI.
- Het moet bekend zijn wat er met de onderzoekdata gebeurt, en in hoeverre dat gebruik ethisch is.
- Er kan een eed/beroepscode worden ingevoerd voor goed AI-gebruik.

Betrouwbaarheid

- De output van AI zal persoonlijk gecheckt moeten worden.
- Het is verplicht de bron te vermelden.
- In elke opleiding zullen ook de beperkingen van genAI duidelijk moeten worden meegenomen.
- Actief bias proberen te vermijden, dus ook opnemen in een peer review. Niet alleen kijken naar wetenschappelijke kwaliteit maar ook naar bias.
- Er moet transparantie zijn over het gebruik van AI. Dit moet de vorm krijgen als een eis aan wetenschappelijke uitgevers.
- Er geldt altijd mens → AI → mens. Dus een check op de data en een check op de uitkomsten.
- Ook creatieve processen vanuit AI moeten herleid worden tot de mensen van wie de eerste ideeën waren.

Kennisverspreiding

- Minder publiceren
- Zorg voor nieuwe vormen van verspreiding van wetenschappelijke kennis voor verschillende doelgroepen.
- Neem als auteur meer de verantwoordelijkheid voor de verspreidingsvormen van je wetenschappelijke artikelen.
- Iedere wetenschapper is verplicht om diens wetenschappelijke output in minimaal drie andere vormen te uiten (een vlog, een populair artikel, etc), daarmee weet je dat beter dat je juist geciteerd wordt.
- Er zou een landelijk platform moeten zijn voor populaire én juiste wetenschap.
- Gebruik AI selectief, zeker in het kader van duurzaamheid. Je kunt denken aan beprijzen of een beperkt aantal slots.
- Goed gedrag op het gebied van AI vergroot je carrièrekansen.

Terugblik en afronding

Aan het einde van de workshop wordt teruggekeken en besproken wat de sessie Aanpak Begeleidingsethiek concreet heeft opgeleverd. Grote gemene deler was dat de aanpak begeleidingsethiek een prettige aanpak is die in korte tijd heeft geleid tot een rijke opbrengst. Een aantal mensen wilde meer ruimte voor discussie. Ook werd er, door het benoemen van de waarden, een snelle convergentieslag gemaakt. Het was een positieve verrassing dat er in de groep eensgezindheid over de kernwaarden was. Er werd opgemerkt dat het jammer was dat big tech zelf niet aan tafel zat, hoewel niet iedereen dat standpunt deelde. Ook lijkt er teveel het gevoel te zijn dat dit ons allemaal overkomt. Maar vanmiddag was een mooi proces waarin we breed gekeken hebben naar dit issue.

De deelnemers

1. **Irene Nooren** programma manager innovatie SURF

2. **Jacco Flipsen** vice-president Medicine and Life Sciences Journals Springer Nature

3. **Hilde van Wijngaarden** directeur Universiteitsbibliotheek Vrije Universiteit

4. **Laurens Hessels** senior onderzoeker Rathenau Instituut, bijzonder hoogleraar RUL, CWTS

5. **Andrea Reyes Elizondo** onderzoeker, CWTS

6. **Martijn Bennis** plv voorzitter Maatschappelijke Coalitie Over Informatie Gesproken

7. **Saskia Haitjema** UHD Diagnostisch Centrum Utrecht UMCU

8. **Daniël Steinmeier** digital preservation officer Koninklijke Bibiliotheek

9. **Rinke Hoekstra** senior director architecture Elsevier

10. **Darco Jansen** manager open science & open access Universiteit van Nederland

11. **Petra Bouts** specialist wetenschappelijke informatie Universiteit Maastricht

12. **Guido Enthoven** oprichter Instituut Maatschappelijke Innovatie

13. **John Doove** teamleider open science SURF

14. **Joost Albers** coördinator research en data science WUR

15. **Bas Haring** hoogleraar publiek begrip van wetenschap RUL

16. **Masoone Shariat** strategic policy advisor for research and innovation University of Twente

17. **Anneke Houdt** netwerkmanager/overheidsliaison Maatschappelijke coalitie Over Informatie Gesproken

15. **Maarten Zeinstra** voorzitter Open Nederland

Moderatoren:
Stephan de Valk en Daniël Tijink (beiden ECP)

aanpak begeleidings ethiek

ECP Platform voor de InformatieSamenleving

Verslag workshop
Aanpak Begeleidingsethiek

Wetenschappelijke kennisverspreiding in het generatieve AI-tijdperk

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