

LES 03

AI Ethiek, Bias, Privacy, Milieu & Soevereiniteit

De diepere verantwoordelijkheden van AI en wat bibliotheken moeten begrijpen, bevragen en verdedigen.

🕒 60–90 min | Alle bibliotheekprofessionals

Waarom deze vragen niet kunnen wachten

AI-implementatie versnelt sneller dan ethische reflectie. Bibliotheken moeten op deze ontwikkeling vooruitlopen.



Ethische mislukkingen gebeuren nu

Gezichtsherkenning heeft onschuldige mensen foutief geïdentificeerd. Wervingsalgoritmen hebben systematisch vrouwen gediscrimineerd. Dit zijn geen hypothetische toekomsten.



Bias is structureel, niet toevallig

AI-bias vloeit voort uit trainingsdata die historische onrechtvaardigheid weerspiegelen. Bias lost zichzelf niet op zonder gerichte interventie.



Gebruikersdata staan onder druk

Elk AI-systeem dat in aanraking komt met vragen van bezoekers, surfgedrag of leesgeschiedenis roept ernstige vragen op rond vertrouwelijkheid. Bibliotheken zijn wettelijk en ethisch verplicht om die te behandelen.

Waarom deze vragen niet kunnen wachten

AI-implementatie versnelt sneller dan ethische reflectie. Bibliotheken moeten op deze ontwikkeling vooruitlopen.



De ecologische kost is reëel

Het trainen van één groot AI-model kan evenveel CO₂ uitstoten als vijf auto's tijdens hun volledige levensduur. Bibliotheken die duurzaamheid belangrijk vinden, moeten dit meewegen.



Wie AI controleert, controleert kennis

Een klein aantal bedrijven controleert de AI-systemen die vandaag bepalen hoe informatie wereldwijd wordt gevonden, gefilterd en gepresenteerd. Dit is een soevereiniteitsvraag.



AI-ethiek

Wat is juist? Wat is schadelijk? Wie beslist?

Ethiek is geen beperking op goede AI; ethiek is er een voorwaarde voor.

Belangrijke kaders voor AI-ethiek

Wereldwijde instellingen hebben kaders ontwikkeld. Bibliotheekmedewerkers moeten ze kennen en toepassen.

UNESCO (2021)

Menselijke waardigheid & inclusie

Het eerste wereldwijde normatieve instrument over AI-ethiek. Legt nadruk op mensenrechten, gendergelijkheid, duurzaamheid en vreedzame samenlevingen. Behandelt expliciet AI in onderwijs en informatie.

EU AI Act (2024)

Regulering gebaseerd op risico

Classificeert AI-systemen volgens risiconiveau: onaanvaardbaar, hoog, beperkt of minimaal. Bibliotheken die AI gebruiken voor profilering van bezoekers of HR-beslissingen kunnen onder hoog-risicocategorieën vallen.

IFLA (2020)

Toegang, privacy en vertrouwen

Verklaring van de International Federation of Library Associations die nadruk legt op open toegang, privacy, vrijheid van surveillance en het recht op toegang tot informatie zonder tussenkomst van een algoritme.

OECD (2019)

Inclusieve groei en verantwoordelijkheid

Vijf principes: inclusieve groei, mensgerichte waarden, transparantie, robuustheid en verantwoordelijkheid. Aangenomen door G20-landen en opgenomen in de meeste nationale AI-strategieën.

Hoe AI schade veroorzaakt: een taxonomie

Verschillende manieren waarop AI-systemen echte schade kunnen veroorzaken.

Type beschadiging	Voorbeeld uit het echte leven	Relevantie voor de bibliotheek
Allocatieve schade	AI weigert iemand een lening op basis van een bevooroordeelde voorspelling.	AI die bibliotheekaccounts markeert op basis van demografische patronen.
Representatieschade	AI koppelt bepaalde groepen aan negatieve kenmerken, vb. “criminele” beelden die raciaal gekleurd zijn.	AI in de catalogus die bij automatische leessuggesties literatuur uit het Globale Zuiden ondervertegenwoordigt.
Kwaliteit-van-dienst-schade	Spraakherkenning werkt slechter voor vrouwen en niet-moedertaalsprekers.	AI-referentietools die beter presteren voor Engelstaligen dan voor anderen.
Interpersoonlijke schade	Deepfakes die worden gebruikt om mensen lastig te vallen, te beschadigen of te manipuleren.	Bezoekers die het doelwit worden van AI-gegenereerde desinformatie op basis van bibliotheekgerelateerde data.
Epistemische schade	AI geeft zelfverzekerd valse informatie, wat vertrouwen in feiten uitholt.	Bezoekers die gegenereerde hallucinaties citeren in academisch of maatschappelijk onderzoek.



Algoritmische bias

Bias is geen bug. Bias zit vaak in de data ingebakken.

Als de wereld onrechtvaardig is en AI leert van de wereld, dan leert AI onrechtvaardigheid.

Waar bias AI-systemen binnenkomt

Bias stapelt zich op in elke fase van de AI-ontwikkelingsketen, niet alleen in de data.

Data-verzameling

Ondervertegenwoordiging van minderheidsgroepen, historische discriminatie die in dossiers is opgeslagen en internetdata die vooral afkomstig zijn uit welvarende, Engelstalige populaties.

Datalabeling

Menselijke annotatoren brengen culturele aannames mee. Labels weerspiegelen het wereldbeeld van wie betaald wordt om te annoteren, meestal niet dat van de onderzochte gemeenschappen.

Model-ontwerp

Optimalisatiematen, zoals accuraatheid voor meerderheidsgroepen, kunnen zwakke prestaties voor minderheden verbergen. Architectuurkeuzes bevatten aannames.

Inzet-context

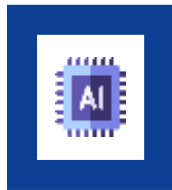
Een tool die in de ene context is getraind en in een andere wordt ingezet, neemt mismatches mee. Een medisch AI-systeem dat op Amerikaanse patiëntendata is getraind, kan falen in Sub-Saharaans Afrika.

Feedback-lussen

Als bevooroordeelde output toekomstige data beïnvloedt, vb. bevooroordeelde wervings-AI → bevooroordeelde personeelsdata → input voor de volgende AI, dan versterkt bias zichzelf na verloop van tijd.

Algoritmische bias

«Een groep golfers»



Algorithmische bias



Algoritmische bias

≡ Translate

English ▼

Turkish ▼

He is a doctor



0 bir doktor



Algoritmische bias

“Garbage in, garbage out” (GIGO) (Rommel in, rommel uit)

Foutieve invoerdata leiden tot foutieve output.





Algorithmische bias

Dutch scandal serves as a warning for Europe over risks of using algorithms

The Dutch tax authority ruined thousands of lives after using an algorithm to spot suspected benefits fraud – and critics say there is little stopping it from happening again.

SCORING OF WELFARE BENEFICIARIES: THE INDECENCY OF CAF'S ALGORITHM NOW UNDENIABLE

27 November 2023



Algoritmische bias

- Amazon ontdekte dat zijn AI-wervingssysteem bevooroordeeld was tegen vrouwen, omdat het was getraind op historische data uit de technologiesector, die overwegend mannelijk was.
- Voorbeelden uit rechtshandhaving tonen foutpercentages die tot 34 procent hoger liggen voor mensen met een donkere huid.
- Medische algoritmen vertoonden bias bij het voorspellen van zorgnoden van zwarte patiënten.
- Advertenties rond huisvesting en werk kunnen algoritmisch worden gestuurd op basis van demografische kenmerken.

Algorithmische bias





Data privacy en surveillance

“Als je niet betaalt voor het product, ben jij het product.”

Data van bibliotheekgebruikers zijn geen grondstof.
Het is een recht dat beschermd moet worden.

Hoe AI-systemen data van bezoekers verzamelen en gebruiken

Datastromen zijn vaak onzichtbaar, maar hun gevolgen niet.

De datastroomketen

- 1 Een bezoeker stelt een vraag via een AI-gestuurde chat of zoekfunctie.
- 2 De vraag wordt verwerkt door een externe AI-API, vaak in een ander land.
- 3 De leverancier logt de vraag om het model te verbeteren.
- 4 Gebruikspatronen worden geaggregeerd en mogelijk verkocht.
- 5 Profielen ontstaan die uitleengedrag, zoek- en vraaggedrag met elkaar verbinden.
- 6 Deze profielen kunnen onder bepaalde wettelijke kaders toegankelijk zijn voor autoriteiten.

Bestaande beschermingsmaatregelen

- ▶ GDPR/AVG in de EU: recht op toegang, verwijdering en dataminimalisatie
- ▶ Veel nationale wetten rond bibliotheekvertrouwelijkheid gelden ook voor digitale interacties
- ▶ IFLA Privacy Framework voor bibliotheeksystemen

Belangrijkste risico's

- ▶ Gegevensoverdracht naar derde landen kan minderwaardige privacybescherming bieden.
- ▶ Een AI-tool kan leveranciers ruime rechten geven op vraagdata.
- ▶ Vragen van bezoekers kunnen gevoelige informatie onthullen, bijvoorbeeld over gezondheid, juridische kwesties of politiek.

Hoe AI-systemen data van bezoekers verzamelen en gebruiken

Vertrouwelijkheid van bezoekers is een wettelijke en ethische basis. AI heft die niet op.

KERNPRINCIPE: de privacybelofte van de bibliotheek stopt niet aan de grens tussen mens en AI. Elk AI-systeem dat gegevens van bezoekers verwerkt, moet aan dezelfde norm voldoen als elk ander bibliotheeksysteem.

1

Voer voor je een AI-tool inzet die gebruikersdata raakt een gedocumenteerde DPIA uit. Beoordeel welke data worden verzameld, waar die naartoe stromen en welke risico's bestaan.

2

Pas dataminimalisatie toe: AI-tools moeten zo worden geconfigureerd dat ze zo weinig mogelijk data verzamelen. Persoonsgegevens moeten waar mogelijk uit AI-prompts worden geweerd.

3

Controleer privacybeleid van digitale leveranciers: lees het volledige privacybeleid en de gebruiksvoorwaarden. Check: worden vragen bewaard? Hoelang? Worden data gebruikt om te trainen? Zijn er subverwerkers?

4

Informeer bezoekers: AI-tools in publieksgerichte diensten, moeten duidelijk worden meegedeeld. Bezoekers moeten AI-diensten kunnen weigeren en toegang hebben tot menselijke alternatieven.

5

Train alle medewerkers rond AI en privacy: elke medewerker moet weten wat wel en niet ingevoerd mag worden. Maak een lijst van toegelaten datatypes en verboden datatypes.



Groepsactiviteit: Data privacy

Cookie-tracking & privacy-instellingen

(zie begeleidersgids)



Milieu-impact

AI heeft een CO₂-voetafdruk. Bibliotheken die duurzaamheid belangrijk vinden, moeten daarmee rekening houden.

De energiekost van “intelligentie” is niet onzichtbaar. Ze zit gewoon verborgen in datacenters.

Milieu-impact

Schaal doet ertoe: de vraag naar grondstoffen door moderne AI-systemen is uitzonderlijk groot.

626.000

kg CO₂-equivalent

De geschatte uitstoot
om één groot
taalmodel vanaf nul te
trainen (MIT-studie)

10×

meer energie

Een ChatGPT-vraag
gebruikt naar
schatting ongeveer 10
keer meer energie dan
een standaard Google-
zoekopdracht

6.4

kubieke meters water

Geschat waterverbruik
van Microsofts AI-
datacenters in 2022
voor koelsystemen

2–4%

*van de wereldwijde
CO₂-uitstoot*

Huidig aandeel van
de ICT-sector, snel
groeiend door de
vraag naar AI.



AI-soevereiniteit en macht

Wie bouwt AI? Wie bezit AI? Wiens waarden weerspiegelt AI?

Controle over AI-infrastructuur is controle over de toekomst van kennis zelf.

Het probleem van de machtsconcentratie van AI

Vijf bedrijven, OpenAI (Microsoft), Google DeepMind, Anthropic (Amazon), Meta AI en xAI, domineren momenteel de meest capabele AI-basismodellen. Hun beslissingen over trainingsdata, toegang, prijszetting, contentmoderatie en eventuele stopzetting bepalen wat miljarden mensen wel en niet kunnen weten.

Concentratie van rekenkracht	Het trainen van grote AI-modellen vraagt enorme rekenkracht, in handen van een beperkt aantal cloudproviders zoals AWS, Azure en GCP
Concentratie van data	De internetdata die worden gebruikt om de meeste LLM's te trainen, zijn vooral afkomstig uit welvarende, Engelstalige landen
Concentratie van arbeid	Contentmoderatie en datalabeling voor AI gebeurt vaak door laagbetaalde werknemers in het Globale Zuiden met minimale arbeidsbescherming
Waardenconcentratie	Beslissingen over wat AI-systemen wel en niet zeggen, tonen of aanbevelen, worden genomen door bedrijfsteams in een klein aantal landen
Reguleringsasymmetrie	AI-regulering ontwikkelt zich het snelst in de EU. Veel andere regio's hebben beperkt toezicht op AI-systemen die hun bevolking beïnvloeden

AI-soevereiniteit: impact voor bibliotheken

Collectiesoevereiniteit: wanneer AI-systemen collecties namens bibliotheken cureren, filteren of rangschikken, verschuiven beslissingen over wat zichtbaar is en wat verborgen blijft van bibliothecarissen naar algoritmen die ontworpen zijn voor commerciële, niet-burgerlijke doelstellingen.

→ *Behoud menselijk toezicht op door AI beïnvloede zichtbaarheid van collecties. Audit rangschikkingen.*

Kennissoevereiniteit: AI-systemen die vooral gebouwd zijn op data uit het Globale Noorden kunnen kennis uit één culturele traditie als universeel voorstellen. Bibliotheken die diverse gemeenschappen bedienen, moeten kritisch evalueren wat AI weet en wat AI uitwist.

→ *Plaats ondervertegenwoordigde kennistradities centraal. Verzet je tegen AI die diversiteit afvlakt.*

Zoeksoevereiniteit: een AI-gestuurde zoekmethode verandert wat bezoekers vinden zonder te veranderen wat bestaat. Een bezoeker die alleen ziet wat AI naar boven haalt, krijgt een verarmde informatieomgeving, gevormd door de training van AI en niet door de werkelijke noden van die bezoeker.

→ *Bied niet-AI-zoekroutes aan. Leer bezoekers gecontroleerde vocabularia gebruiken.*

Juridische soevereiniteit: data die door AI-leveranciers worden verwerkt, kunnen onder buitenlandse surveillancewetgeving vallen. Bibliotheken moeten ervoor zorgen dat grensoverschrijdend AI-gebruik voldoet aan nationale privacyregels.

Vraag juridisch advies over verplichtingen rond dataresidentie vóór AI-aankoop of aanbesteding.

Groepsactiviteit: de ethische audit

Groepen van 4 tot 6 personen | 25 minuten | Kies één AI-tool die je bibliotheek gebruikt of overweegt.

Ethiek

Heeft deze tool een gepubliceerd ethisch kader?
Wie heeft dat opgesteld?
Welke faalscenario's en schadevormen zijn gedocumenteerd?

Bias

Is de tool onafhankelijk getest op demografische bias?
Welke talen en gemeenschappen zijn ondervertegenwoordigd in de trainingsdata?

Privacy

Welke data worden verzameld en opgeslagen en voor hoelang?
Is de tool GDPR/AVG-conform? Is er documentatie?

Milieu

Publiceert de leverancier gegevens over energieverbruik of uitstoot?
Draaien de datacenters op hernieuwbare energie?

Soevereiniteit

In welke landen worden de data verwerkt?
Weerspiegelt het contentbeleid van de leverancier de waarden en rechten van onze gemeenschap?