nota

TER ONDERTEKENING

Aan: MOCW

Emancipatie

Van

[redacted]

Datum
22 januari 2024

Referentie
42922374

Afgestemd met

D.Kennis

Bijlagen
4

Aanbieding onderzoeksrapport Tweede Kamer:
Kunstmatige Intelligentie en lhbti+ emancipatie

Aanleiding

Het kennisinstituut Movisie heeft in opdracht van OCW een eerste verkenning uitgevoerd naar hoe lhbti+-emancipatie en kunstmatige intelligentie (AI) zich tot elkaar verhouden (bijlage 1). Dit onderzoeksrapport wordt zonder beleidsreactie door u aan de Tweede Kamer aangeboden, omdat momenteel wordt gekeken op welke wijze de inhoud van het rapport kan worden gebruikt binnen het emancipatiebeleid.

Geadviseerd besluit

U wordt geadviseerd de bijgevoegde kamerbrief te ondertekenen.

Toelichting

Algemeen

- Movisie heeft een eerste verkenning uitgevoerd naar hoe lhbti+-emancipatie en kunstmatige intelligentie (AI) zich tot elkaar verhouden.
- Dit is onderzocht door middel van expertinterviews en een literatuurverkenning in vier maatschappelijke deelgebieden: onderwijs, arbeidsmarkt, gezondheidszorg en (online) veiligheid.
- De komende periode zal door DE gekeken worden hoe deze uitkomsten op verschillende deelterreinen van het emancipatie beleid gebruikt kunnen worden. Hierover wordt de Tweede Kamer bij volgende gelegenheid geïnformeerd.
- Daarnaast is o.a. de directie Kennis betrokken bij het (interdepartementale) traject rondom de aanstaande AI act vanuit de Europese Uni. Een aantal van de zorgen en risico's met betrekking tot mensenrechten die in dit rapport gesignaleerd worden, worden in hun algemeenheid (los van lhbti+ emancipatie) ook in het huidige voorstel meegenomen. Indien relevant, kunnen de uitkomsten van het rapport meegenomen worden in het vervolg traject. De Tweede Kamer wordt hierover geïnformeerd via kabinetsbrede brieven over het algemene AI beleid.

Opbrengsten samenvatting

- Er is nog weinig onderzoek naar en kennis over de relatie tussen lhbti en AI;
- Uit de ***literatuurverkenning*** komt naar voren dat AI veelvuldig wordt ingezet op de vier onderzochte terreinen, maar dat de inclusiviteit van de data waarop AI-systemen worden getraind niet voldoen.

    - Er liggen bijvoorbeeld kansen voor het ondersteunen van onderwijsprofessionals met hun administratieve taken, waardoor meer tijd voor het geven van onderwijs overblijft (thema onderwijs). Ook het inrichten van werving en selectie vrij van discriminatie of onbewuste vooroordelen (thema arbeidsmarkt) is een mooie kans.
    - Tegelijkertijd liggen er risico's op de loer, in algemene zin door vooroordelen in de trainingsdata, waardoor systemen in elk domein alsnog kunnen discrimineren.
    - En meer specifiek is bijvoorbeeld het gebrek aan toegankelijke informatie over lhbtiq+ personen doordat algoritmes dit foutief classificeren als porno (thema onderwijs) een risico.
    - Een ander risico is het moeten delen van gevoelige informatie met onveilige websites, omdat een ander systeem daarvan afhankelijk is (thema veiligheid).
- De resultaten per deelgebied staan vermeld in bijlage 1 onder nota.
- Experts uiten in een ***expertsessie*** hun zorgen over de vertrouwelijkheid, transparantie, representativiteit, opslag, heteronormativiteit en de mensenrechtentoets van bestaande datasets. De resultaten van de expertinterviews staan per deelgebied vermeld in bijlage 2 onder de nota.
    - Binnen de risico's wordt door de experts gewezen op de gebrekkige dataverzameling die tot discriminerende algoritmes kan leiden. Gebrek aan (juiste) data leidt mogelijk tot onterecht "computer says no" besluiten. Ook noemen zij het gebrek aan bescherming tegen online geweld en de onduidelijkheid over het opslaan en gebruiken van gegevens over lhbtiq+ personen. Deze kunnen in strijd zijn met het recht op privacy. Bijvoorbeeld door een niet beveiligde opmerking in een dossier van een lhbtiq+ persoon te plaatsen over diens seksuele oriëntatie, op een plek deze dan leesbaar is ook op momenten waarop dit gegeven niet relevant is.
    - De kansen verschillen per thema. In het thema onderwijs ligt bijvoorbeeld een kans voor meer inclusieve content voor docenten, terwijl op het thema arbeidsmarkt ook de potentie ligt om daders van pesten op de werkvloer op te sporen. Binnen het thema gezondheidszorg biedt het ontbreken van kennis volgens de expert een uitgelezen mogelijkheid om de zorg inclusiever te maken door deze groep meer inspraak te geven, dat kan leiden tot betere zorg. Binnen het thema veiligheid ligt er de kans dat door online discriminerende opmerkingen de omvang van het probleem meetbaar wordt.

- Experts onderstrepen de noodzaak van onderzoek en een multi-stakeholder aanpak om zo tot beleid, innovaties en handelingsperspectieven te komen die inclusief zijn.

Informatie die niet openbaar gemaakt kan worden

N.v.t.

Bijlage 1: Resultaten literatuurverkenning

1. Onderwijs
   - In het onderwijs komen steeds vaker toepassingen van AI voor zoals bij het overbrengen en toetsen van kennis en bij het informeren van docenten over presentaties van leerlingen.
   - Risico's:
     - Beperkte toegang tot onderwijs door gebruik van algoritmen bij toelating;
     - Beperkte toegang tot informatie over lhbti+ door gebruik van filters op schoolcomputers;
     - Ontoereikende mogelijkheden voor het ondersteunen van leerlingen door binair onderscheid jongen-meisje.
   - Kansen:
     - Ondersteunen van onderwijsprofessionals door overnemen administratieve taken;
     - Interactie met menselijk ingrijpen: onderwijsprofessionals kunnen gebruikmaken van bestaande infrastructuur van het GSA-netwerk om hun ondersteuning aan lhbtiq+ leerlingen te optimaliseren.

   Met name op het gebied van ondersteuning van lhbti+ leerlingen liggen er mogelijkheden voor de inzet of verdere toepassing van AI.

2. Arbeidsmarkt
   - AI kan worden toegepast in het transparanter maken van wervings– en selectieprocessen en het opsporen van discriminatie in vacatureteksten zoals SZW al doet.
   - Risico's:
     - (On)bedoelde impact op de al minder gunstige arbeidsmarktpositie van lhbti+ personen door automatisering;
     - AI werkt nog niet vlekkeloos in het wervings- en selectieproces door onbewuste vooroordelen van de recruiters die als input worden gebruikt voor het algoritme;
     - Mogelijke discriminatie in geautomatiseerde online vacatures, die onderscheid maken op basis van leeftijd, sekse of afkomst;
     - Mogelijke nadelen van het inzetten van video-interviewing, waar het nog onbekend is in hoeverre de trainingsdata representatief zijn voor lhbti+ personen.
   - Kansen:
     - Betere inrichting van het wervings– en selectiesysteem zodat ze minder of niet discrimineren, zodat werkgevers of recruiters geen onbewust vooroordeel kunnen meenemen in hun keuze.

3. Gezondheidszorg
   - De gezondheidszorg is één van de belangrijkste domeinen waar AI haar meerwaarde heeft laten zien zoals nauwkeuriger diagnoses stellen, behandelingsplannen voor patiënten vereenvoudigen en medicijnen sneller en goedkoper te ontwerpen.
   - Risico's:
     - Gebrek aan transparantie en uitleg van medische beslissingen;
     - Beperkte beschikbaarheid van kwalitatief goede gegevens, bij lhbti+ personen is er vaak sprake van *oversampling*, waardoor er goed inzicht in de diversiteit en complexiteit ontbreekt;

    - Mogelijk verankeren en vergroten bestaande ongelijkheid, aangezien kwetsbare groepen niet altijd goed vertegenwoordigd zijn in trainingsdatasets.
- Kansen: aanzienlijke kansen om de zorg en resultaten voor patiënten te verbeteren kosten te verlagen en de gezondheid van de bevolking te beïnvloeden.

4. <u>(Online) Veiligheid</u>
    - AI kan worden ingezet om lhbti+ personen online te ondersteunen, door verbinding te maken en ervaringen te delen, maar ook veel kans tot haatspraak.
    - Risico's:
        - Inbreuk op het recht op gelijke behandeling door onderscheid tussen groepen personen in het domein van sociale zekerheid;
        - Niet wenselijk om gebruik te maken van modellen en checklists, omdat de persoonlijke identiteit uniek is;
        - Delen van gevoelige informatie op mogelijk onveilige platforms en afhankelijkheid daarvan.
    - Kansen:
        - Benutten van online accounts van lhbti+ personen voor dataverzameling;
        - Mogelijkheid tot online ondersteuning;
        - Beter inzicht in maatschappelijke houding ten aanzien van lhbtii+ personen en thema's.

Bijlage 2: Resultaten expertinterviews

Algemene risico's

- Representativiteit van data, ontwikkelaars en besluitvormers: in de praktijk blijken deze zelf niet representatief voor de lhbti+ gemeenschap, met als risico dat oplossingen voor vraagstukken met een heteronormatieve bril worden bekeken;
- Transparantie, vertrouwelijkheid en ethiek: in de praktijk tonen deze datasets vaak een overwicht aan heteroseksuele personen.
- Commerciële partijen zetten AI in om het gedrag van gebruikers te sturen, (online) veiligheid is dan van ondergeschikt belang voor deze partijen.

1. Onderwijs
   - Risico's:
     - Binariteit van het geslacht in de datasets die worden verzameld om de prestaties van leerlingen te meten en monitoren, niet in lijn met moderne tijdgewricht waar deze binariteit steeds meer ter discussie staat.
     - Onduidelijkheid over het opslaan van data en het effect van gedragssturing: de toepassingen van AI kunnen op verschillende wijze inbreuk maken op het privacy recht.
   - Kansen:
     - Voorbereiden op een arbeidsmarkt die vergeven is van AI;
     - Betere en inclusievere content door docenten;
     - Aandacht vragen voor normen en grensoverschrijdend gedrag binnen docententeams;
     - Inclusievere opleidingen kunnen een aantrekkingskracht uitoefenen op lhbti+ personen, specifiek op kansen voor werving van leerlingen als een betere cultuur binnen opleidingen.
     - Vergroten van toegankelijkheid van onderwijs.

2. Arbeidsmarkt
   - Risico's:
     - Algoritmen die voor recruitmentsoftware dienen het doel dat mensen die qua profiel lijken op de trainingsdata aan te nemen. In veel gevallen zijn dit voornamelijk witte mannen.
     - Wervings – en selectieprocedures bestaan grotendeels uit het maken van testvragen en persoonlijkheidstests die als doel hebben om personen te werven met vergelijkbare profielen als degene in managementposities, dit zijn voornamelijk witte, hoogopgeleide mannen.
   - Kansen:
     - Individuen kunnen gebruik maken van trucs om online wervingssystemen te omzeilen;
     - Kan een rol spelen in het opsporen van potentiële daders van pesten op de werkvloer.

3. Gezondheidzorg
   - Risico's:
     - Gebrek aan representatie in de datasets;
     - Ethische aspecten aan dataverzameling (onbeveiligde data op verschillende plekken, gebruik van het gegeven wanneer dat niet

relevant is, het is een bijzonder persoonsgegeven zonder de benodigde grondslag of opslagmethode).
    - Inclusie van genderidentiteit is van ondergeschikt medisch belang, omdat er nadruk ligt op de biologische sekse-identiteit.
- Kansen:
    - Ontbrekende kennis is een uitgelezen kans om algoritmen te ontwikkelen die gezondheidsprofessionals in hun werk kunnen ondersteunen.
    - Opeisen van het recht van spreken en bemoeienis door lhbti+ personen bij gesprekken over AI en data om inclusieve datasets te bouwen.

4. <u>(Online) Veiligheid</u>
    - Risico's:
        - Heteronormativiteit in het ontwerp van AI systemen en het soms moedwillig uitsluiten van leden van de lhbti+ gemeenschap.
        - Veiligheid wordt bijna nooit gewaarborgd.
    - Kansen:
        - AI maakt het mogelijk om gelijkgestemden te vinden die je minder makkelijk in de samenleving zou kunnen vinden.
        - Opzetten en onderhouden van veilige ruimten die je kan monitoren en beschermen.
        - Inzet van sociale media en algoritmen bij de lhbti+ emancipatie.
        - Zicht krijgen op houdingen ten opzichte van lhbti+ personen in verschillende online en offline omgevingen en opsporen van potentiële daders van pesten van lhbti+ personen.