AI en de Bibliotheek: Zeven Principes

Mei 2020

KB nationale bibliotheek

AI en de Bibliotheek

AI sinds een aantal jaren werken we in de Koninklijke Bibliotheek (KB) met Artificiële Intelligentie (AI). We onderzoeken de toepassingen en relevantie daarvan voor de bibliotheek. Ook hebben we een aantal *Proof of Concepts* uitgevoerd. We verwachten dit jaar de eerste AI-toepassing operationeel te maken. AI is hierbij geen doel op zich. We zien AI als een vorm van automatisering. Weliswaar een zeer geavanceerde vorm, maar uiteindelijk toch automatisering.

AI is een relevant hulpmiddel voor de digitale transformatie die de bibliotheken op dit momemt doormaken. Daar kun je vanuit vier invalshoeken naar kijken:

1. Onze klanten: Hoe verandert AI de dienstverlening van de bibliotheek?
2. Ons werk: Hoe verandert AI de manier waarop we werken?
3. De AI zelf: Wat kan de bibliotheek bijdragen aan de ontwikkeling van AI?
4. De maatschappij: Hoe kan de bibliotheek helpen AI op verantwoorde wijze in te zetten?

Bij de eerste twee zijn we gebruikers van AI; bij de laatste twee spelen we een rol in de ontwikkeling van AI als vakgebied. Dankij onze eeuwenlange ervaring met het collectioneren, cureren, organiseren en toegankelijk maken van informatie zijn wij als bibliotheek bij uitstek de partij die AI-toepassingen met onze datacollecties kan trainen en vormen.

Zeven Principes

Van de nationale bibliotheek van Nederland mag verwacht worden dat zij een positie inneemt over de principiële, maatschappelijke vraagstukken die AI met zich meebrengt. Omdat we een inherent waardesysteem hebben, ethische vraagstukken al decennialang adresseren, kunnen we een bijdrage leveren aan de ‘opvoeding’ van AI.

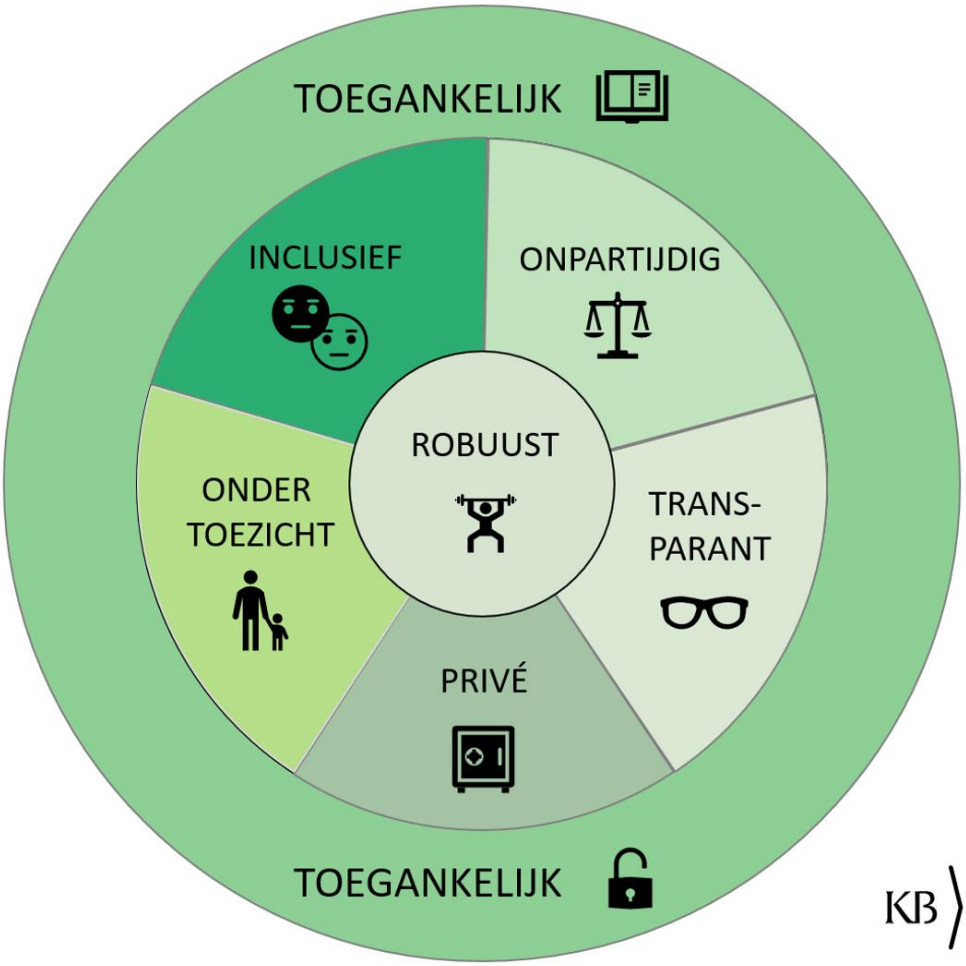
Met dat als uitgangspunt heeft de KB het initiatief genomen om principes te formuleren waaraan we AI-toepassingen kunnen toetsen.

Bij het formuleren van deze zeven AI-principes zijn we uitgegaan van de vraag hoe AI ons kan helpen de opdracht uit te voeren die de Nederlandse samenleving ons heeft gegeven. Die luidt dat we zorg dragen voor het geschreven woord en iedereen in staat stellen om te lezen, te leren en onderzoek te doen. We delen daarmee het centrale uitgangspunt van het *AI for good* initiatief van de VN, dat AI kan en moet worden ingezet om de wereld beter te maken. Wat een betere wereld is, hebben de VN geformuleerd in 17 Sustainable Development Goals (SDG). Vele, uiteenlopende initiatieven en projecten worden door deze SDG’s gestuurd en ook de *AI for good* initiatieven worden eraan getoetst.

In de visuele weergave van de samenhang tussen de zeven principes gebruiken we drie concentrische cirkels, waarvan de middelste uit vijf segmenten bestaat. Om de cirkel heen bevinden zich de gebruikers; de klanten van de bibliotheken. Het vergroten van de toegankelijkheid van informatie ligt op de buitenste cirkel en dus het dichtst bij de gebruikers, waarmee de positieve mogelijkheden van AI benadrukt worden.

Door robuustheid in de binnenste cirkel te plaatsen, geven we aan dat elk AI-systeem in de kern een systeem is waaraan dezelfde betrouwbaarheidseisen gesteld worden als aan elk ander softwaresysteem.

De principes in de middelste ring adresseren vijf ethische kwesties die door de toepassing van AI op maatschappelijk of individueel niveau relevant worden. Makers en gebruikers van AI-systemen moeten op die kwesties een antwoord formuleren.



KB nationale bibliotheek

1. Toegankelijk

De bibliotheek zet AI primair in voor het toegankelijk maken van informatie voor het publiek en voor het bevorderen van (digitale) geletterdheid van alle burgers.

We zetten AI in voor de missie van de KB: Nederland slimmer, creatiever en vaardiger maken. Door publicaties met behulp van AI-toepassingen beter en slimmer toegankelijk te maken en te houden, nu en in de toekomst, stellen we iedereen in staat om te lezen, te leren en onderzoek te doen.

Daarmee dragen we bij aan de duurzame ontwikkelingsdoelstellingen (SDG) van de VN, met name SDG 16.10 *publieke toegang bieden tot informatie* en SDG 4.6 *geletterdheid en rekenvaardigheid bevorderen*.

Met behulp van AI kan de leesvaardigheid van mensen verbeterd worden met een op de persoon afgestemd leren-lezen-programma.

Een andere toepassing is AI die het mogelijk maakt om in gesproken, natuurlijke taal vragen te stellen aan een AI-systeem dat alle boeken, kranten en tijdschriften gelezen heeft. Hiermee wordt informatie laagdrempelig en voor iedereen bereikbaar.

Een laatste voorbeeld is het door AI laten voorlezen van teksten aan mensen die het lezen (nog) niet beheersen. AI brengt de normaal gesproken dure productie van luisterboeken binnen handbereik door geschreven teksten automatisch in gesproken woord om te zetten.

2. Robuust

We werken alleen met AI-toepassingen die robuust ontworpen en ontwikkeld worden en die betrouwbaar zijn in het gebruik.

AI-applicaties zijn software-applicaties. De robuustheid daarvan wordt bepaald door de kwaliteit van ontwerp, realisatie, testen, beheer en documentatie. AI leveranciers dienen gebruik maken van bewezen methoden, standaarden, normen, frameworks en certificeringen voor software engineering, aangevuld met nieuwe standaarden voor machine learning algoritmen en datasets voor training en testing.

Een AI-toepassing moet kwalitatief aan hoge normen voldoen. Een AI-toepassing kan niet de ene keer antwoord A geven, en een volgende keer antwoord B, als er niets veranderd is dat relevant is voor de uitkomst.

Het is misschien overkomelijk als aanbevelingssoftware voor boeken vijf suggesties doet waarvan er vier kloppen. Maar als dezelfde AI-software gebruikt wordt voor het aanraden van wetenschappelijke artikelen en die vijfde, gemiste aanbeveling bevat juist de laatste stand van onderzoek, dan wordt het problematisch.

3. Inclusief

We ontwikkelen en gebruiken alleen datasets en AI-toepassingen die inclusief zijn.

We realiseren ons dat alle AI inherent kwetsbaar is voor *bias* (vooringenomenheid).

Als KB stellen we ten behoeve van onderzoek naar en ontwikkeling van AI datasets beschikbaar die zo min mogelijk *biased* zijn ten opzichte van leeftijd, etniciteit, religie, genderidentiteit, seksuele oriëntatie, herkomst en politieke voorkeur.

De praktijk is echter hard: feitelijk bestaan er ook in de (digitale en fysieke) magazijnen van de KB geen datasets die 100% unbiased zijn, al was het maar door veranderende opvattingen over collectiebeleid door de jaren heen.

Belangrijker dan bias voorkomen, is dan ook om te weten waar en in welke mate er bias optreedt, zodat die weggenomen of gecompenseerd kan worden, waardoor inclusiviteit gewaarborgd blijft.

4. Onpartijdig

We ontwikkelen en gebruiken geen AI-toepassingen die actief het gedrag of denken van mensen beogen te manipuleren.

Hoewel AI-systemen zijn ontworpen om mensen te ondersteunen bij het nemen van beslissingen en het maken van keuzes, mag dit nooit misbruikt worden door gebruikers ongevraagd in richtingen te sturen die ze zonder deze AI-toepassing ook niet zouden kiezen.

Partijdigheid bij commerciële of politieke keuzes en beslissingen doet bovendien afbreuk aan de diversiteit. Net als inclusiviteit kan de onpartijdigheid van AI-systemen worden beïnvloed door gerichte (trainings)data en algoritmen.

De bibliotheek dient hier haar rol als neutrale expert op te pakken en haar gebruikers te begeleiden en van onpartijdig advies te dienen.

5. Onder toezicht

We ontwikkelen en gebruiken alleen AI-toepassingen waar op cruciale punten menselijk toezicht is. 'No human no AI.'

AI dient ter ondersteuning van menselijk activiteiten maar moet niet als zelfstandig beslissend systeem optreden. AI kan vele taken zelfstandig uitvoeren maar moet uiteindelijk onder toezicht van mensen staan, die het eerste en het laatste woord hebben.

Menselijk toezicht hoeft niet per se betrekking te hebben op het trainen van AI of het controleren van elke output van het systeem. Een van de doelen van AI als technologie is immers om dit overbodig te maken. Het gaat om het kritisch testen en beoordelen van AI-toepassing, zowel op het gebied van de hier geformuleerde principes, als op terreinen waarop AI nog lang niet volwassen is, zoals het leggen van oorzakelijke verbanden en het tonen van sociale intelligentie.

6. Transparant

We ontwikkelen en gebruiken waar mogelijk alleen AI-toepassingen waarvan de algoritmes, trainingsdata en -methode transparant zijn.

AI hoeft geen 'black box' te zijn waarvan zelfs de ontwerper niet kan uitleggen waarom deze AI tot een bepaalde beslissing is gekomen. Transparantie kan geboden worden door toepassing van inzichten uit *Explainable AI* (XAI). Een belangrijk uitgangspunt van XAI is het 'recht op een verklaring', dat is het recht dat een individu heeft om uitgelegd te krijgen hoe een AI-systeem tot een bepaalde conclusie is gekomen, vooral als die financiële, juridische of sociale gevolgen heeft.

XAI ontwikkelt hiertoe methoden en technieken die het menselijke experts mogelijk maken om de uitkomsten van een AI-toepassing te kunnen verklaren, ook als volledige transparantie door complexiteit of vertrouwelijkheid van de algoritme niet altijd mogelijk is.

7. Privé

We ontwikkelen en gebruiken alleen AI-toepassingen waarvan we zeker zijn dat die de persoonlijke levenssfeer van onze medewerkers en gebruikers respecteren.

Dit principe is niet exclusief op AI van toepassing. Er is uitgebreide wetgeving rond privacy en de persoonlijke levenssfeer, met name in de context van IT-systemen, waaraan ook elke ontwerper en gebruiker van AI-systemen zich dient te houden.

De reden om privé toch als afzonderlijk principe te formuleren, ligt in de maatschappelijke zorg die er over privacy in relatie tot AI bestaat. Van gezichtsherkenning van mensen die zich anoniem in de openbare ruimte denken te bewegen, tot het combineren van gegevens uit verschillende, op zichzelf al bedreigende *big userdata* verzamelingen ten behoeve van commercieel doelen: de verontruste reacties hierop zijn begrijpelijk en terecht.

De rol van de bilbiotheek is hier niet om deze zorgen te relativeren, maar in tegendeel om een *safe haven* te zijn, een plek waar alles wat privé is gerespecteerd wordt, en beschermd.

-.-

mei 2020
KB, Nationale Bibliotheek
Jan Willem van Wessel
janwillem.vanwessel@kb.nl