

7 principes om kunstmatige intelligentie in te zetten

De Koninklijke Bibliotheek wil een positie innemen over de principiële, maatschappelijke vraagstukken die Artificiële Intelligentie (AI) met zich meebrengt. Daarom heeft de KB zeven principes geformuleerd waaraan zij haar werk met én voor AI kan toetsen.



Jan Willem van Wessel

Hoofd sector
Verwerking en Behoud
van de Koninklijke
Bibliotheek

Al sinds een aantal jaren worden in de Koninklijke Bibliotheek stappen gezet op het gebied van Artificiële Intelligentie (AI). AI is daarbij geen doel op zich, het is een vorm van automatisering. Weliswaar een vergevorderde vorm, maar uiteindelijk toch automatisering.

AI is een relevant hulpmiddel voor de digitale transformatie waar we middenin zitten. Hierbij zijn vier invalshoeken te onderscheiden:

Onze klanten: hoe verandert AI de dienstverlening van de bibliotheek?

Ons werk: hoe verandert AI ons voortbrengingsproces?

De AI zelf: wat kan de bibliotheek bijdragen aan de ontwikkeling van AI?

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 pakken we onze rol in de ontwikkeling, zo je wilt zelfs in de 'opvoeding' van AI als vakgebied. Met 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 data kan trainen en vormen.

Positie innemen

Van de nationale bibliotheek van Nederland mag bovendien verwacht worden dat zij een positie inneemt over de principiële, maatschappelijke vraagstukken die AI met zich meebrengt. Omdat we een inherent waardesysteem

hebben en 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 zeven principes te formuleren waaraan we ons werk *met* AI en *voor* AI kunnen toetsen. Bij het formuleren van deze AI-principes gaan we uit van de vraag hoe het ons kan helpen de missie uit te voeren die de Nederlandse samenleving ons heeft opgedragen. Die luidt dat we zorg dragen voor het geschreven woord en iedereen in staat stellen om te lezen, te leren en onderzoek te doen. Zo dragen we eraan bij dat Nederland slimmer, vaardiger en creatiever wordt.

Sustainable Development Goals

We delen daarmee het centrale uitgangspunt van het *AI for good*-initiatief van de VN, namelijk dat AI kan en moet worden ingezet om de wereld beter te maken. Wat een betere wereld is, heeft de VN geformuleerd in zeventien Sustainable Development Goals (SDG; zie tinyurl.com/yafaeg9p). Vele uiteenlopende initiatieven en projecten worden door deze SDG's gestuurd en ook de *AI for good*-initiatieven (aiforgood.itu.int) worden eraan getoetst.

Reageer!

De zeven principes zijn nog in ontwikkeling. In de loop van dit jaar wil de KB deze afronden. Heb je opmerkingen, aanvullingen of suggesties? Stuur ze naar Jan Willem. vanWessel@kb.nl.

Toegankelijk

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 ook 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*.

Robuust

We gebruiken alleen AI-toepassingen die robuust zijn. Dat wil vooral zeggen: betrouwbaar, compleet en getest. Zowel in technisch opzicht als in termen van de uitvoer. Een AI-toepassing kan niet in 95 procent van de gevallen werken; het moet werken volgens vooraf geformuleerde verwachtingen. Een AI-toepassing kan niet de ene keer antwoord A geven, en een volgende keer antwoord B, als er niets veranderd

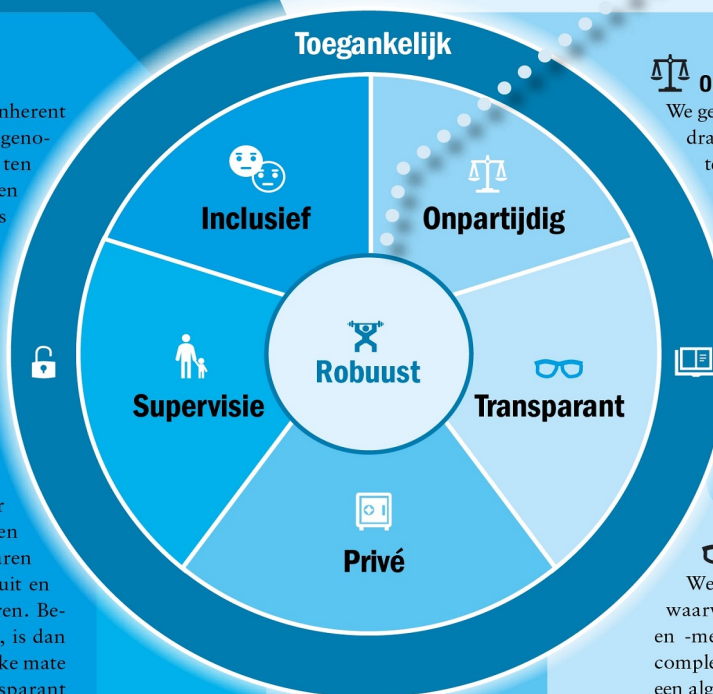
is dat relevant is voor de uitkomst. Het is misschien niet erg als aanbevelingssoftware voor boeken vijf suggesties doet waarvan er vier kloppen en de vijfde niet. Maar als dezelfde AI-engine gebruikt wordt voor het aanraden van wetenschappelijke artikelen en die vijfde, gemiste aanbeveling bevat juist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

Inclusief

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 honderd procent unbiased zijn, al was het maar door veranderende opvattingen over collectiebeleid door de jaren heen en door de missie 'alles' uit en over Nederland te collectioneren. Belangrijker dan bias voorkomen, is dan ook om te weten waar en in welke mate er bias optreedt, en daar transparant over te zijn. We helpen onze gebruikers door ze hierop te wijzen en onze expertise in te zetten om ze daar op de juiste manier mee te laten omgaan.

Supervisie

No human no AI. We gebruiken AI ter ondersteuning van menselijk activiteiten, niet als zelfstandig beslissend systeem. AI kan vele taken zelfstandig uitvoeren maar zal uiteindelijk altijd onder supervisie van mensen staan, die het laatste woord hebben. Dat hoeft niet per se het trainen van AI of het controleren van de resultaten te zijn. Het kan ook gaan om het testen en beoordelen van een AI-toepassing, voordat deze in gebruik genomen wordt.



Onpartijdig

We gebruiken geen AI die actief het gedrag of denken van mensen beoogt te manipuleren. Hoewel AI-systemen zijn ontworpen om mensen te ondersteunen bij het nemen van beslissingen en keuzes, mag dit nooit misbruikt worden door gebruikers in richtingen te sturen die ze zonder deze AI-toepassing ook niet zouden kiezen. Partijdigheid kan bovendien afbreuk doen aan de diversiteit van organisaties, diensten en producten.

Transparant

We gebruiken waar mogelijk AI waarvan de algoritmes, trainingsdata en -methode transparant zijn. Waar complexiteit of vertrouwelijkheid van een algoritme volledig inzicht in de weg staan, dienen de makers/leveranciers bekend en vertrouwd te zijn. We volgen de recente ontwikkelingen rond *Explainable AI* (XAI) en willen daar een actieve rol in spelen.

Privé

We gebruiken alleen AI waarvan we zeker zijn dat die de persoonlijke levenssfeer van onze medewerkers en gebruikers in acht neemt. We gebruiken alleen AI die overeenstemt met de voor ons geldende nationale en internationale wet- en regelgeving.