

Kunstmatige Intelligentie

→ Richting 'Franken-modellen'

→ Efficiëntere aanpak van AI-systemen

→ Nieuwe manieren om toegang te krijgen tot gegevens

→ Richting betrouwbare AI

→ Toegankelijke computing en modellen



Kunstmatige intelligentie

Kunstmatige intelligentie (AI: Artificial Intelligence) en, meer specifiek, machine-leren (ML) hebbende de laatste jaren hun vleugels uitgeslagen. Dit heeft geleid tot indrukwekkende prestaties, zoals het verslaan van 's werelds beste Go-speler met AlphaGO en bijna levensechte beeldgeneratie door systemen zoals Dall-E. Het leidde echter ook tot zorgen en discussies over de mogelijke risico's en schade die AI-systemen kunnen veroorzaken.

AI-technologie heeft al een sterke invloed op de onderzoeks- en onderwijswereld, zowel wat inhoud betreft, als de werking, en dit zal ongetwijfeld in de toekomst zo blijven. AI is een gewild studiegebied bij onderzoekers en ML is een waardevolle onderzoeksmethode.

Daardoor neemt het aantal toepassingen de laatste tijd toe. Binnen het onderwijs is AI minder ver ontwikkeld voor wat betreft de implementatie ervan, zoals te zien is in de trendmanifestaties die hier worden gepresenteerd. De belofte van onderwijsapparaten met AI (AIED's: Artificial Intelligence Educational Devices) groeit echter, zowel in het klaslokaal (microniveau), als in instellingen (mesoniveau) en de samenleving in het algemeen (macroniveau).

AI wordt vaak ingezet binnen een complex systeem dat de technologische infrastructuur binnen sociaal-technische contexten bevat. Deze ontwikkeling wordt ook geïllustreerd door trends op het gebied van bijvoorbeeld berekeningen, data en hosting van diensten.

Het wordt ook weerspiegeld in trends met betrekking tot manieren waarop deze technologische elementen in specifieke community's ingezet zou kunnen worden om een betrouwbaar ecosysteem te vormen.



TREND #1

Richting 'Franken- modellen'

Publieke waarden

	Autonomie	Onafhankelijkheid van onderwijs
	Rechtvaardigheid	Gelijkheid Integriteit Betrouwbaarheid van informatie Transparantie Democratisch bestuur Duurzaamheid Inclusiviteit
	Menselijkheid	Veiligheid

Gereedheid

VOLG

PLAN

DOE

Drijfveren

#Internationalisering #Dataïsme
#Decentralisatie #Energiezuinigheid #Open Science
#Open source-ontwikkeling

Vroege modellen voor ML zijn ontworpen voor een specifieke taak. Recente ontwikkelingen tonen aan dat we nu complexere modellen kunnen bouwen die verschillende soorten input combineren in een systeem met bredere capaciteiten. Innovaties in architectuur en multimodale, multi-objectieve training van modellen maken dit mogelijk. Modellen worden getraind met verschillende soorten informatie en een verscheidenheid aan taken. Instellingen met toegang tot die modellen hebben een concurrentievoordeel. Deze (Franken-)modellen worden vaak getraind met zwakke labels of zelfs zonder toezicht. De methodologie heeft betrekking op het gebruik van zelftoezicht, wat leidt tot generieke capaciteiten. Deze modellen kunnen worden gezien als een nieuw Zwitsers zakmes, dat nieuwe capaciteiten en mogelijkheden voor automatisering ontsluit, wat leidt tot waardecreatie.



Funderingsmodellen

De term 'funderingsmodel' is bedacht door onderzoekers van het Stanford Institute for Human-Centred Artificial Intelligence voor het vastleggen van de toename in grote modellen die op grote schaal zijn getraind en zich kunnen aanpassen aan een breed scala aan taken. Deze modellen bieden alle een platform van capaciteiten die later kunnen worden aangepast aan specifieke toepassingen, zoals BERT, chatGPT-4, chatGPT, DALL-E en stabiele diffusie.

 open voorbeeld



DeepMind's Flamingo-model

Flamingo is een visueel taalmodel (afbeeldingen + tekst) ontwikkeld door DeepMind. Het combineert een visueel en een taalmodel, wat betekent dat het een breed scala aan taken met betrekking tot die modaliteiten kan aanpakken. Flamingo is ook een 'few-shot'-leermodel, wat betekent dat het nieuwe taken kan leren met slechts een paar extra inputs. De volgende stap is het overschakelen van tekst-naar-afbeelding naar van tekst-naar-videotoepassingen. Deze technologie biedt potentie om te helpen bij het creëren van betrokken persoonlijke assistenten voor de onderwijssector.

 open voorbeeld

 open voorbeeld

IMPACT

Deze modellen kunnen nieuwe mogelijkheden ontsluiten voor zowel onderzoek als onderwijs: informatie halen uit ingewikkelde systemen, nieuwe complexe data genereren, nieuwe datamodaliteiten toevoegen aan

bestaande pijplijnen, bestaande taken automatiseren/versnellen en meer interactieve en boeiende cursussen ontwerpen. Deze modellen beschikken over end-to-end-mogelijkheden, waardoor ze complexe problemen kunnen modelleren. Zoals vaak het geval is, zijn die modellen gevoelig voor problemen met vertekening, vooroordelen en auteursrechten. Niet-uniforme ingebruikname van deze instrumenten kan ook de hiaten in efficiëntie tussen de actoren vergroten. Het aansturen van de getrainde modellen wordt steeds belangrijker om integriteit te waarborgen. Er bestaat een potentieel risico dat modellen te krachtig worden en onze menselijke waarden gaan bedreigen. We verwachten dat dergelijke generieke, grote, voorgetrainde modellen het uitgangspunt zullen worden voor toekomstig taakontwerp en training.

Multimodaal leren in de gezondheidszorg


De multimodale aanpak helpt onderzoekers om kennis en waarde uit een exascale dataset te halen zonder dat daar grote gelabelde, geannoteerde gegevens voor nodig zijn. Het toezicht wordt uitgevoerd door gebruik te maken van getrainde netwerken om labels te leveren die op hun beurt door een actieve leerinterface worden gecorrigeerd. We kunnen beelden van weefsels combineren met de bijbehorende diagnostische tekst. Een vergelijkbare impact is te verwachten op gebieden waar kennisextractie afhankelijk is van basiscapaciteiten uit funderingsmodellen, zoals het combineren van data uit sensors met meerdere modaliteiten of het combineren van resources in een leeromgeving.

 open voorbeeld

TREND #2

Efficiëntere aanpak van AI-systemen

Publieke waarden

 **Autonomie** Onafhankelijkheid van onderwijs

 **Rechtvaardigheid** Integriteit | Duurzaamheid | Effectiviteit

 **Menselijkheid**

Gereedheid

VOLG

PLAN

DOE

Drijfveren

#Internationalisering #Dataïsme
#Decentralisatie #Energiezuinigheid #Open Science #Open source-ontwikkeling

Naarmate het gebruik van AI-systemen toeneemt, zien we een trend richting efficiënte training en implementatie van AI-systemen. Deze trend staat gelijk aan een inspanning om zowel de data- als de berekeningskosten van training en implementatie van AI-systemen te verlagen. Door economische en ecologische gevolgen is het niet altijd haalbaar om steeds grotere modellen te trainen in het kader van economische kosten als duurzaamheid. Op modelniveau gebruikt inductieve vertekening kennis van het bestudeerde systeem bij voorbaat om de modelarchitectuur te beperken. Een opmerkelijke klasse van opkomende netwerken is het gepulst neurale netwerk ('Spiking Neural Networks'), dat belooft neuronefficiënter te zijn dan een klassiek kunstmatig neuraal netwerk, d.w.z. dat het complexer gedrag kan bereiken met minder neuronen dan traditionele modellen.

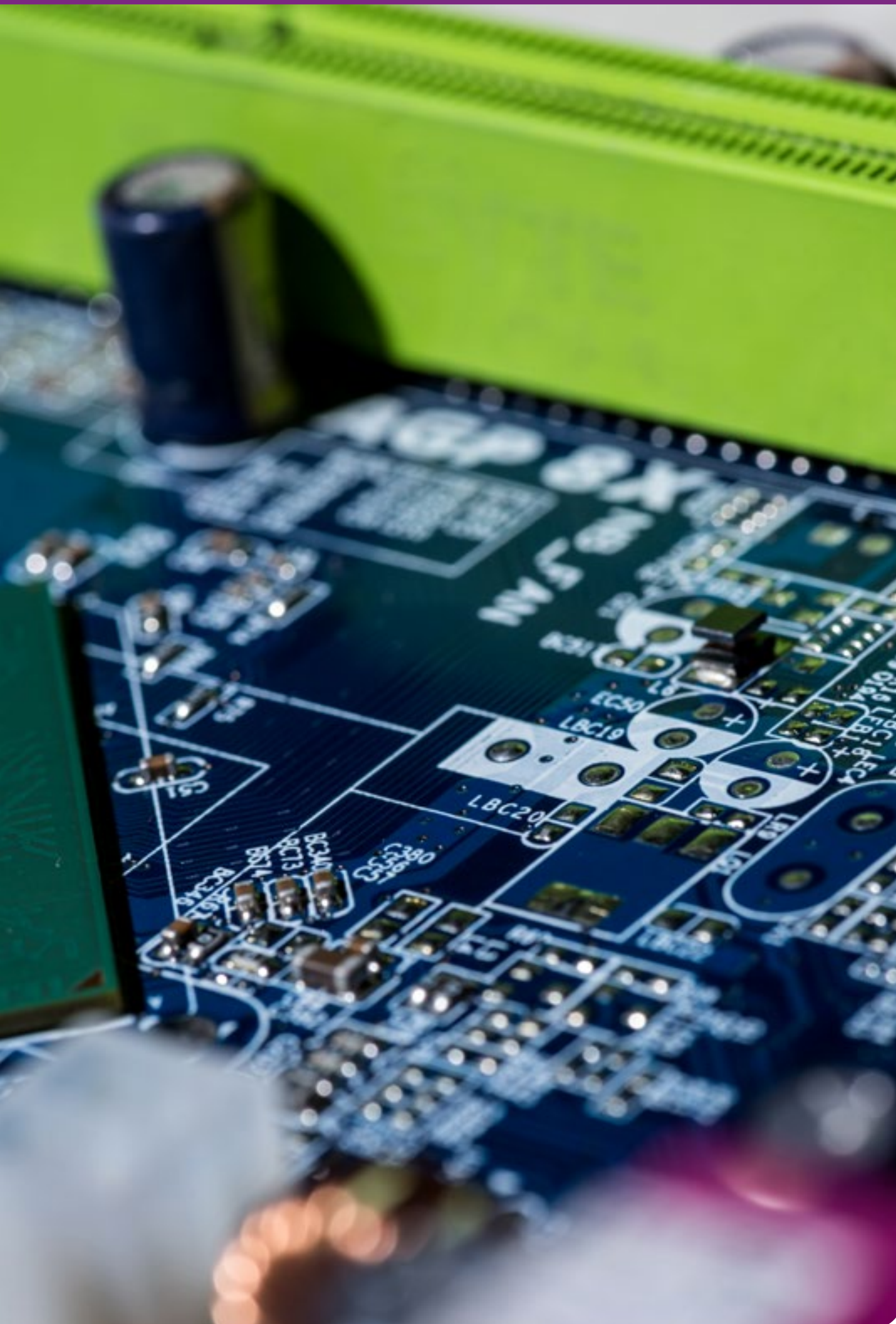
Dit wordt gedaan met behulp van biologisch geïnspireerde activeringen. Naast het verminderen van de complexiteit van data en modellering, treedt er ook een verschuiving op naar AI-specifieke hardware om modellen effectief te trainen en uit te voeren. De overgang van CPU naar GPU voor bepaalde computingtaken heeft geleid tot de ontwikkeling van nog specifiekere hardware.

Natuurkunde-geïnformeerde neurale netwerken

Bij meerdere projecten wordt rekening gehouden met fysieke wetten (een vorm van inductieve vertekening) bij het ontwerpen van de simulatiestroom om de gegevensbehoefte en modelcomplexiteit te beperken. Op die manier verwerpen de modellen automatisch alle mogelijke 'niet-fysieke' oplossingen voor het probleem, waardoor het aantal resultaten en dus de trainingstijd drastisch afneemt.

 open voorbeeld

 open voorbeeld



Grafische neurale netwerken

GNN zijn een ander voorbeeld van inductieve vertekening. Deze netwerken zijn met name doeltreffend voor taken waarbij gegevens afkomstig zijn uit niet-Euclidische ruimte (d.w.z. waarbij de fysieke ruimte irrelevant is). Het meest voor de hand liggende voorbeeld bestaat natuurlijk uit sociale netwerken, maar deze netwerken kunnen worden toegepast in een verrassend breed scala aan domeinen (bijv. natuurlijke taalverwerking, scheikunde).

 [open voorbeeld](#)

In de richting van AI-specifieke hardware

Met de opkomst van AI zijn er ook nieuwe mogelijkheden ontstaan voor alternatieve rekenkundige- en gegevensarchitecturen om te profiteren van de specifieke activiteiten die nodig zijn voor deep learning-algoritmen. Sinds de vroege TPU van Google ontwikkelen veel bedrijven nu innovatieve oplossingen om algoritmen efficiënt te trainen en te implementeren (bijv. Graphcore, Cerebras, Habana, Rain Neuromorphics).

 [open voorbeeld](#)

 [open voorbeeld](#)




IMPACT

Effectievere AI-systemen zullen de invoering van AI verhogen door ze toegankelijker te maken voor een breder gebruikerspubliek. Inductieve vertekening zal waarschijnlijk een aanzienlijke invloed hebben op onderzoek naar AI en met AI. Bovendien zal de ontwikkeling van AI-specifieke hardware de mogelijkheden van AI-systemen opnieuw bevorderen, en deze mogelijkheden zullen de vraag naar nieuwe infrastructuren doen toenemen. De belangrijkste drijfveer hier is om zoveel mogelijk uit beperkte middelen te halen. Hiervoor zijn echter zeer gespecialiseerde hardware en vaardigheden nodig.

TREND #3

Nieuwe manieren om toegang te krijgen tot gegevens

Publieke waarden

-  **Autonomie** Onafhankelijkheid van onderwijs
-  **Rechtvaardigheid** Integriteit, gelijke kansen | effectiviteit
-  **Menselijkheid**

Gereedheid

VOLG

PLAN

DOE

Drijfveren

#Privacy #Europese wetgeving #Digitale economie #Data-economie #Datagovernance #Automatisering

Het centraliseren en delen van data is niet altijd haalbaar vanwege de complexiteit van de organisatie, buitensporige kosten of privacyproblemen. Er worden steeds meer oplossingen ontwikkeld om toegang te krijgen tot inzichten uit moeilijk te delen data. Dit gebeurt ofwel door gedistribueerde training (gefedereerd leren), training met synthetische gegevens, ofwel door het delen van beveiligde gegevens en computeromgevingen. AVG en andere initiatieven maken dataprivacy tot een topprioriteit, waardoor deze trend wordt versterkt. De hoge (data)kosten voor het van het begin af aan trainen van nieuwe modellen is ook een stimulans om op zoek te gaan naar creatieve manieren om met minder data te werken. Deze trend helpt organisaties om controle en eigenaarschap van gegevens te behouden en kan helpen om de ecologische impact van het AI-ecosysteem te verminderen.



Gefedereerd leren

In gebieden waar bepaalde data of een deel van de data niet kan worden gedeeld (bijv. gezondheidszorg en onderwijs), wordt het cruciaal om algoritmen te trainen zonder volledige toegang tot de gegevens te hebben. Dit is wat gefedereerd leren oplost door een netwerk van knooppunten te ordenen, waarbij elk knooppunt gedeeltelijk een algoritme leert met alleen zijn data, voordat het model centraal wordt samengevoegd, zonder ooit toegang te hebben gehad tot de volledige dataset zelf.

 [open voorbeeld](#)



ODISSEI Beveiligde supercomputer

Via een beveiligde supercomputing-omgeving worden zeer privacygevoelige data van het CBS op een veilige manier toegankelijk gemaakt voor onderzoekers. Dit democratiseert de toegang tot bevolkingsinformatie en versnelt het onderzoek binnen de sociale wetenschappen.

 open voorbeeld

 open voorbeeld

IMPACT

Deze trend is met name nuttig om AI mogelijkheden te bieden in gebieden die tot nu toe niet in staat zijn geweest om volledig gebruik te maken van AI, omdat toegang tot gegevens lastig is (bijv. gezondheidszorg, onderwijs). Door het delen van en de toegang

tot data te vergemakkelijken, kunnen nieuwe onderzoeksinzichten worden verkregen die tot nieuwe mogelijkheden leiden. De privacyrisico's in verband met de synthetische data-benadering mogen echter niet worden onderschat. Zodra de gepaste infrastructuur en protocollen zijn ontwikkeld, kunnen synthetische data helpen om de privacy te beschermen en kunnen er modellen worden getraind op basis van grotere datasets. Hierdoor ontstaan er ook meer mogelijkheden voor internationale onderzoekssamenwerkingen. Synthetische data bieden de eigenaars van de data ook meer controle.

Synthetische gegevens

Een andere benadering van dit probleem is het genereren van een reeks synthetische data op basis van een model van de werkelijke verdeling. Dit is krachtig, maar werkt alleen als er in de eerste plaats al een krachtig basismodel voor het genereren van synthetische data kan worden getraind. Als er een synthetische dataset kan worden gebruikt, kunnen toepassingen gebruik maken van een gelijkwaardige dataset, maar zonder privacyproblemen.

 open voorbeeld

TREND #4

Richting betrouwbare AI

Publieke waarden

-  **Autonomie** Onafhankelijkheid van onderwijs | Bescherming van het privéleven en persoonsgegevens
-  **Rechtvaardigheid** Gelijkheid | Transparantie | Democratisch bestuur
-  **Menselijkheid** Veiligheid

Gereedheid

VOLG

PLAN

DOE

Drijfveren

#Europese wetgeving #Verantwoorde technologieën #Privacy #CO2-voetafdruk #Globalisering #Digitale economie #Data-governance #Diversiteit #Inclusie #Gelijkheid #Digitale geletterdheid #Gendergelijkheid

De toegenomen introductie en prestaties van AI-modellen hebben geleid tot belangrijkere maatschappelijke vraagstukken op het gebied van verantwoordelijkheid, vertrouwen en transparantie. Dit blijkt uit de focus van de Europese Commissie op betrouwbare AI en toekomstige AI-wetgeving. Binnen betrouwbare AI komen verschillende maatschappelijke en technologische ontwikkelingen samen om deze maatschappelijke vraagstukken aan te pakken. Kaders en wetgeving vormen een basis voor de vereisten waaraan AI moet voldoen. Een sterkere focus op betrouwbaarheid, normen en interoperabiliteit in ontwikkeling leidt tot robuustere en transparantere AI-systemen. Democratisering van AI leidt tot meer hulpbronnen om van te leren en meer normen om te volgen, wat ook nieuwkomers helpt om aan de slag te gaan.



Ethische AI-richtlijnen

Binnen een groeiende discours over het verantwoorde gebruik van AI bespreken academici, NGO's en bedrijven de impact van AI en de ethische implicaties ervan. Deze discussies omvatten onderwerpen als vertrouwen, transparantie, eerlijkheid en verantwoordelijkheid. Er zijn veel hulpmiddelen naar voren gekomen, maar één van de belangrijkste is de ethische richtlijnen die zijn uitgebracht door de deskundigengroep voor AI van de Europese Commissie.

 [open voorbeeld](#)



Standaardisering en professionalisering

Van dataverzameling tot training en implementatie, de typische engineeringwerkstroom is complex. Net als bij best practices en tools voor software-engineering is er een tendens gaande in AI, onder leiding van grote bedrijven en onderzoeksgroepen, om dergelijke werkstromen door bijvoorbeeld MLOps mogelijk te maken. Standaardisering van activiteiten en implementatie leidt tot een meer volwassen, professionele omgeving. Dit is cruciaal voor interoperabiliteit, maar helpt ook bij transparantie om vertrouwen te bevorderen en vertekeningen te voorkomen.



Algoritmeregisters

Om het gebruik van algoritmen transparant te maken, zijn meerdere gemeenten en overheidsinstellingen begonnen met het gebruik van algoritmeregisters om de samenleving inzicht te geven in het gebruik ervan. In de toekomst zal dit waarschijnlijk verplicht worden voor alle overheidsorganisaties. Door openbaar te delen waar, wanneer en hoe algoritmen worden gebruikt, kan transparantie en verantwoordelijkheid worden geboden aan relevante stakeholders, zoals burgers, gebruikers van toepassingen, de media en de relevante autoriteiten.



Door de community geleide initiatieven

Grote AI-modellen zijn vaak alleen toegankelijk voor grote onderzoekslaboratoria en bedrijven. Community geleide initiatieven zoals BLOOM zijn erop gericht AI-modellen te democratiseren door ze tegen minder middelen beschikbaar te stellen aan onderzoekers in kleinere laboratoria. Deze initiatieven hebben ook betrekking op meer partijen in de trainingsfase om te zorgen voor betere transparantie en robuustere modellen. BLOOM is een samenwerking van meer dan 1000 onderzoekers uit meer dan 70 landen, die een groot taalmodel hebben getraind voor 46 natuurlijke talen en 13 programmeringstalen.



IMPACT

Betrouwbare AI is een cruciale drijfveer voor alle partijen die aan AI werken, zowel in onderzoek als in het onderwijs. Het formaliseren van goede werkwijzen en normen draagt bij aan het professionaliseren van de community en levert hulpbronnen voor verantwoord gebruik en snelle, praktische hulpmiddelen. Voldoende aandacht voor betrouwbare AI kan voorkomen dat (1) schade wordt toegebracht aan community's door onverantwoord gebruik van AI, en (2) community's worden opgesplitst met verschillende normen voor het werken met AI-methoden.

TREND #5

Toegankelijke computing en modellen

Publieke waarden

-  **Autonomie** Vrijheid van onderzoek en onderwijs
-  **Rechtvaardigheid** Gelijkheid | Inclusiviteit
-  **Menselijkheid**

Gereedheid

VOLG PLAN DOE

Drijfveren

#Onderzoeksomgeving #Digitale geletterdheid
#Dataïsme #Decentralisatie #Automatisering
#Connectiviteit #Digitale economie
#Globalisering

AI-methoden groeien uit tot wetenschappelijk instrument dat eenvoudig gebruikt kan worden. Als ‘nieuw hulpmiddel’ tussen andere hulpmiddelen moet AI toegankelijk zijn als basisproduct tegen lage kosten. Aangezien de meeste implementaties plaatsvinden in cloud-omgevingen, wordt AI steeds toegankelijker in het openbare domein. Het gebruik van abstractie-lagen (zoals autoML- of MLaaS-benaderingen) helpt om directe toegang te bieden tot complexe computationele vaardigheden voor niet-technische deskundige stakeholders. Dit leidt echter tot een afweging tussen toegankelijkheid en de risico's als afhankelijkheid van aanbieder en de privacykwesties bij het gebruik van deze applicaties in de commerciële cloud-infrastructuur.



Low-code/no-code

Met meer geavanceerde softwarelagen bovenop de computerinfrastructuur vindt er een verschuiving plaats naar low-code/no-code. Verschillende servicetoepassingen maken complexe taken mogelijk met weinig of geen programmering. Terwijl klassieke HPC-centra nog steeds een voorsprong hebben op het gebied van geavanceerd intensief onderzoek, maakt low-code/no-code rekenkracht beter toegankelijk voor potentieel nieuwe gebruikers.

 open voorbeeld



Naar machineren als service (MLaaS)

In verband met het voorgaande voorbeeld is de groei van platforms om het delen van datasets, voorgetrainde modellen en het gemak om voort te bouwen op bestaande modellen een goed voorbeeld van de voortdurende democratisering van AI-methoden en de lagere instapkosten. Deze platforms helpen ook bij de transparantie en standaardisering van methoden.

 open voorbeeld

Modelontwerp met autoML mogelijk maken

Een belangrijk onderdeel in de huidige werkstroom van een data scientist is het experimenteren met modelarchitectuur voor een bepaald probleem en het afstemmen van de trainingsprocedure om de beste resultaten te behalen. Dit is zeer tijdrovend en sterk afhankelijk van de ervaring van de wetenschapper. In plaats daarvan stelt autoML voor om dit deel van de werkstroom te automatiseren en taken met meer toegevoegde waarde aan de gebruiker over te laten. Zoals hoe het probleem moet worden gekaderd, welke maatstaf relevant is en hoe resultaten op een verantwoorde manier kunnen worden geïnterpreteerd.

 open voorbeeld

IMPACT

Deze trend is een mes dat aan twee kanten snijdt: het democratiseert de toegang tot rekenfaciliteiten en AI-methoden, maar ten koste van potentiële afhankelijkheid van één aanbieder, privacyproblemen en onverantwoord gebruik van die algoritmen door een gebrek aan diepgaande kennis. Onderzoeks- en onderwijsinstellingen moeten richtlijnen voor verantwoord gebruik van AI leren, volgen en inzicht krijgen in de impliciete afwegingen bij het kiezen van verschillende computingplatforms. De democratisering van AI-methoden zou moeten helpen bij het boeken van vooruitgang op gebieden die tot nu toe buiten beschouwing werden gelaten.

Meer over kunstmatige intelligentie

Contact

Matthieu Laneuville

Programmamanager AI
matthieu.laneuville@surf.nl

Damian Podareanu

*Teamleider high performance
Machine Learning*
damian.podareanu@surf.nl

Duuk Baten

Adviseur voor Verantwoord AI
duuk.baten@surf.nl

Bertine van Deyzen

Projectmanager voor AI in het onderwijs
bertine.vandeyzen@surf.nl

Meer info

