

De ethiek rond Artificial Intelligence en Machine Learning: waar staan we nu en wat kan beter?

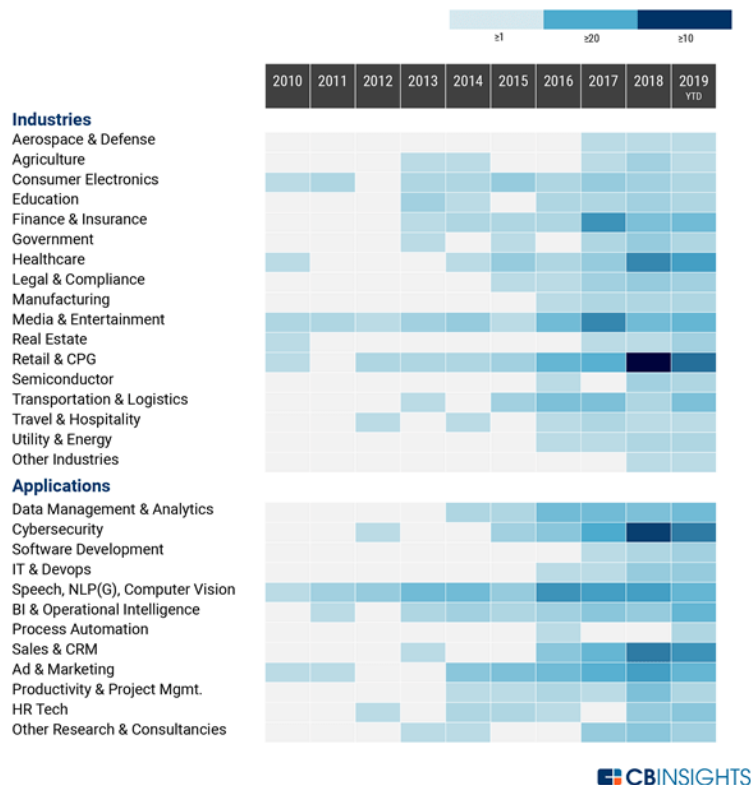
Mensen willen vaak vooroplopen. Hoe kunnen we nóg beter worden? Wat wordt de volgende trend? Afgelopen jaren is duidelijk geworden dat [Artificial Intelligence \(AI\)](#) 'the next big thing' is, specifiek [Machine Learning \(ML\)](#). Dit komt vooral door de toegenomen proceskracht en computervaardigheden. Maar Artificial Intelligence en Machine Learning hebben ook valkuilen.

Hergebruik van Artificial Intelligence

Artificial Intelligence is een veelvoorkomend buzzwoord dat door allerlei organisaties [1] gebruikt wordt: van corporate tot ngo en van start-up tot overheid. Iedereen wil 'iets met AI', maar helaas is er geen algemeen geaccepteerde definitie van Artificial Intelligence. De enige overeenkomst tussen alle definities is dat er allerlei computertechnieken en ingewikkelde processen bij komen kijken.

Het doel van al deze technieken en processen is intelligente machines nog slimmer maken, waardoor ze steeds beter worden in bijvoorbeeld patroonherkenning, computervisie en taalverwerking [2]. Dit is ook het overkoepelende doel van Artificial Intelligence. Maar zodra er nieuwe technologieën ontstaan, verandert ook wat wel en niet onder AI valt. Wat ooit innovatief was, wordt normaal en is na verloop van tijd geen kunstmatige intelligentie meer.

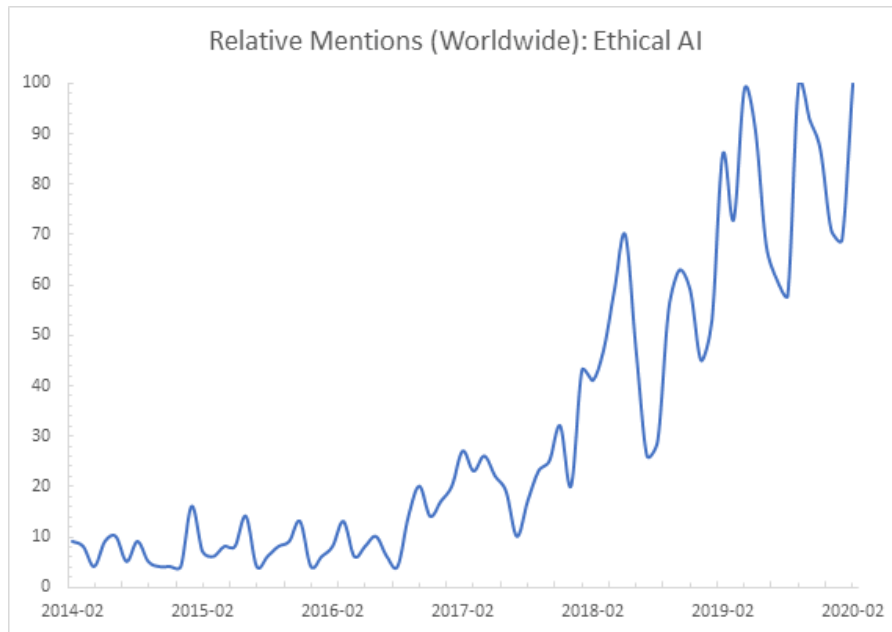
HEATMAP: CONCENTRATION OF AI ACQUISITIONS BY CATEGORY (2011-2019 YTD)



Figuur 1: Grafiek van het aantal AI-gerelateerde aanvragen per sector

Ups en downs

Als je kijkt naar wat Artificial Intelligence organisaties te bieden heeft, is het lastig om hier volledig objectief naar te kijken. De beloftes van AI zijn enorm. AI-systemen presteren soms zelfs beter dan mensen, bijvoorbeeld in complexe vakgebieden (medische diagnostiek). Hierdoor kunnen organisaties meer bereiken met minder budget. Bovendien zorgt AI vaak voor veel nieuwe mogelijkheden binnen de organisatie.



Figuur 2: het relatief aantal keer dat op 'Ethical AI' werd gezocht in Google

Daarom willen organisaties graag – en het liefst als eerste – Artificial Intelligence integreren in hun business. De attitude is 'adapt-or-die'. Maar ook bij AI komen uitdagingen kijken, onder andere vanwege de ontbrekende definitie en de toenemende vraag naar technologische innovaties. Elke nieuwe trend krijgt vroeg of laat met het volgende vraagstuk te maken: hoe zit het eigenlijk met de ethische verantwoording van dit product/deze dienst?

Wat gebeurt er met mijn persoonsgegevens?

Ook dit is een veelvoorkomend vraagstuk tegenwoordig. Het is geen verrassing dat AI-systemen die volledig gebaseerd zijn op enorme hoeveelheden persoonlijke data, impact hebben op de privacy van burgers. De vraag is wat die impact precies inhoudt. Een veelgebruikt acroniem (een leidraad) hiervoor is Fairness, Accountability, en Transparency in Machine Learning (FAT/ML). Dit acroniem gaat over ethiek, de introductie ervan in AI en 'black box' algoritmes. Bovendien garandeert dit acroniem anti-discriminatie (Fairness), een eerlijk proces (Accountability) en volledig begrip bij beide partijen (Transparency).

Tips en tricks voor ethisch verantwoorde kunstmatige intelligentie

Helaas staat ethisch verantwoorde AI nog in de kinderschoenen, dus we moeten hier kritisch naar kijken. Wel zijn er een paar punten die je in gedachten kunt houden als je Artificial Intelligence zo ethisch verantwoord mogelijk wilt maken. Niet alleen binnen je team, maar ook binnen je hele organisatie.

1. Neem 'echte' data scientisten aan

Er wordt van alles verwacht en geëist van potentiële data scientisten in allerlei sectoren. Daarom wordt de term 'data scientist' gebruikt voor de meest uiteenlopende functies [4]. Zorg ervoor dat de mensen die je aanneemt ook daadwerkelijk de taken kunnen uitvoeren die je ze toeschrijft. Zorg er ook voor dat je als organisatie gebruikmaakt van een ethische standaard als je werkt met machine learning modellen en data van je klanten.

2. Begrijp het verschil tussen correlatie en oorzakelijkheid

Machine learning modellen voorspellen op basis van de correlaties en patronen die uit de data voortkomen. Hierdoor is het wel lastig vast te stellen of een model om kan gaan met veranderende data. Een causale gevolgtrekking onderzoekt deze mogelijke oorzaak-gevolgrelatie en bekijkt wat er gebeurd kan zijn toen het model te weinig data had.

Bij machine learning ligt de focus meer op voorspellen (met correlatiepatronen) en minder op oorzakelijkheid. Daarom is het belangrijk om te onthouden dat correlatie en causaliteit van elkaar verschillen. Dit verschil beperkt de conclusie(s) die je kunt trekken uit de resultaten van je model. Een voorbeeld: postcodes worden vaak gebruikt als variabele voor ethische achtergrond. Als jouw model deze informatie gebruikt om beslissingen te nemen, kan dit leiden tot een bias (vooroordeel) voor verschillende etnische groepen.

3. Houd rekening met vooroordelen

Data scientisten moeten garanderen dat het systeem of model dat zij ontwerpen, rekening houdt met mogelijke vooroordelen uit het verleden. Deze biases vloeien vaak voort uit beleid, economische keuzes en gedrag. Het is dus niet de bedoeling dat een model zich ontwikkelt op basis van in het verleden veroordeelde situaties.

Als je rekening houdt met vooroordelen, dan wordt je model robuuster en representatiever. Bijvoorbeeld: in casus X waren alle mensen die een bepaalde moord toegeven, aanwezig bij vergadering Y. Als er geen causale link is gevonden tussen de moord en de aanwezigheid bij de vergadering, dan is het verstandig om deze informatie niet te gebruiken in het model. Het kan namelijk zo zijn dat een persoon die niet aanwezig was bij vergadering Y, ook schuldig is. En deze persoon moet natuurlijk niet uitgesloten worden door het model.

4. Geef uitleg en wees transparant

Wees duidelijk over mogelijke beperkingen en nadelen. Zo weet je zeker dat iedereen dezelfde verwachting heeft. Deze heldere communicatie is essentieel als de uitkomsten of voorspellingen van het algoritme door andere individuen worden geïmplementeerd dan degene(n) die het algoritme ontworpen hebben (en dit is vaak het geval). Het is belangrijk om hier een protocol voor te schrijven en in ieder geval te bespreken hoe de uitkomsten geïmplementeerd en gebruikt moeten worden.

5. Ontwikkel of implementeer een auditplatform

De bovenstaande punten kunnen ingewikkeld lijken als er geen concreet framework is dat je kunt aanhouden. Daarom is de ontwikkeling of implementatie van een auditplatform handig. Dan heb je een zekere standaard om modellen te certificeren. Dit gebeurt vanuit zowel technisch, communicatief als uniform perspectief. Een voorbeeld: Totta data lab heeft – samen met [Verdonck, Klooster & Associates \(VKA\)](#) – een [algoritmekeurmerk](#) ontworpen. Dit is een audittool die de betrouwbaarheid en effectiviteit van algoritmes meet.

6. Houd je vakgebied in de gaten

Data science ontwikkelt zich in een razend tempo, en de gebruikers hiervan moeten meegaan met de laatste ontwikkelingen. Dit houdt ook in dat je jezelf (en anderen) regelmatig moet opleiden en nieuwe mogelijkheden moet zien te vinden in dit vakgebied. Daarnaast is het belangrijk om de voortgang op het gebied van causaliteit en verklaarbaarheid in de gaten te houden. Dit leidt vaak tot een betere transparantie tussen data scientists, hun modellen en de klanten die dit model gebruiken. Zo hebben wij onlangs een open AI-versie van ons fraudemodel ontworpen binnen het sociale zekerheidssysteem. Dit algoritme geeft meer inzicht in de beslisprocedures van het model, zodat je precies kunt zien op basis waarvan het model keuzes maakt.

Wees alert (en ethisch verantwoord)

Machine Learning en Artificial Intelligence worden gebruikt in projecten die een blijvende impact hebben. De bovenstaande punten zorgen ervoor dat dit op een ethisch verantwoorde manier kan gebeuren. Houd wel in gedachten dat er geen eenduidig antwoord is op de vraag wat 'ethisch verantwoorde kunstmatige intelligentie' is. Wat wel altijd opgaat, is dat je kritisch moet denken, open moet staan voor nieuwe technieken en intensief moet samenwerken met betrokken afdelingen en stakeholders. Daarnaast kun je de bovenstaande punten in gedachten houden. Maar het belangrijkste is: keep your mind open!

Referenties

[1] Zie figuur 1.

[2] Sutart J. Russel and Peter Norvig, Artificial Intelligence: A Modern Approach (1995).

[3] Pamela McCorduck, Machines who think: A personal inquiry into the history and prospects of Artificial Intelligence (2004).

[4] Den Daas, S. (2019, 27 mei). Wat Zoek Jij In Een Data Scientist? [blog]. URL: <https://www.tottadatalab.nl/2019/05/27/wat-zoek-jij-data-scientist/>.