De Amsterdamse televisiezender AT5 onderschept toeristen die door de Piet Heintunnel zijn gefietst. "Gestuurd door de navigatie van Apple," aldus de verbolgen toerist.

Computer says yes

Maar begrijpen wij de vraag wel?

'Het was Apple Maps die ons deze kant opstuurde. Als de politie iets van ons wil, klaag ik Apple aan', jammerde een toerist tegen een verslaggever van het Amsterdamse AT5 nadat hij door de Piet Heintunnel was gefietst. 'We hadden wel dood kunnen zijn'. Niet computer says no, maar computer says yes, bleek het probleem. Kunstmatige intelligentie die mensen aanzet om fouten te maken. Dat was niet wat ons beloofd was.

Door **Ruben Boyd**
Beeld **Redactie**

Problemen als deze zullen vaker voorkomen in een cultuur waarin we blind vertrouwen op onze handige digitale hulpjes. Waarom zouden wij het beter weten dan de kunstmatige intelligentie die vliegensvlug ogenschijnlijk perfecte routes uit de mouw schudt? Maar ons vertrouwen blijkt soms fataal. In Amerika kwam een voetganger om het leven na een aanrijding met een zelfrijdende Uber. Uit het onderzoeksrapport bleek dat de Uber de voetganger wel op tijd had gezien, maar dat de 'automatisch remmen'-functie was uitgeschakeld omdat er al meerdere keren vals alarm was geweest. En 'zien' is misschien te veel gezegd. Het slimme systeem categoriseerde de onfortuinlijke voetganger binnen de laatste secondes respectievelijk als 'onbekend object', 'voertuig' en 'fietser', met bijbehorende verschillen in voorspelde trajectoriën. De bestuurder keek door YouTube aanbevolen filmpjes op haar slimme telefoon en besefte te laat wat er gebeurde.

Wanneer de zelflerende algoritmes van bedrijven in de fout gaan of suboptimale adviezen geven, kunnen mensen 1) hun gezonde verstand gebruiken, 2) het advies uitvoeren en erachter komen dat het fout was, of 3) ook dat niet doorhebben. Tot dusver hebben die fouten, naast tragische auto-ongelukken, nog niet tot heel grote problemen geleidt. Weinig mensen maakt het iets uit wanneer een slim aanbevelingssysteem op basis van koopgedrag en zoekgeschiedenis een advertentie aanbiedt die de plank misslaat. Jij klikt er niet op, het algoritme registreert dat het niet tot de gewenste actie heeft geleid, en leert hiervan door zijn parameters aan te passen. En wanneer je de aanbevelingen van een bepaald platform niet nuttig of handig vindt, kun je ervoor kiezen een ander platform te gebruiken. De zelflerende algoritmes selecteren op basis van de verzamelde data de beste, nuttigste of waarschijnlijkste opties, maar het individu heeft in theorie nog steeds het laatste woord.

Algoritmes zouden ons nu ook moeten verlossen van de grotere problemen die de wereld kent en waar wij, met alleen menselijk intelligentie, schijnbaar tekortschieten. Zelflerende systemen zouden met een combinatie van veel data en slimme algoritmes overheidsdienstverlening kunnen optimaliseren, epidemieën voorspellen en criminelen herkennen in een menigte. Gezondheidszorg en onderwijs kunnen door personalisatie beter aansluiten op de wensen en behoeftes van het individu. Alles wordt bijgehouden en alles wordt data. Maar hoe komen we daar? Op digitaleoverheid.nl is de kop 'kunstmatige intelligentie' vooralsnog akelig leeg.

Katten en honden

Veel van de rooskleurige vooruitzichten komen voort uit het succes van deep learning, een methode van zelflerende algoritmes die gebaseerd zijn

Computer says no!

Op maandag 17 september vond in Den Haag de rondetafeldiscussie 'Computer says no!' plaats, georganiseerd door iBestuur en Pegasystems.

Jaap van den Herik, hoogleraar Informatica en Recht, Universiteit Leiden, wil een debat opzetten tussen de 'socially driven', bijvoorbeeld juristen, en de 'technology driven', zoals data scientists. "Met een multidisciplinaire samenwerking en dito samenleving als uitgangspunt."

Marlies van Eck, zelfstandig adviseur AI en recht, vindt dat wat geldt voor de analoge wereld ook geldt voor de digitale wereld: het beginsel fair play, het zorgvuldigheidsbeginsel, het verdedigingsbeginsel en het gelijkheidsbeginsel. De overheid is nog lang niet klaar voor AI. "Eerst de rechtsbescherming op orde, dan pas verder!"

Het ministerie van JenV inventariseerde de verschillende soorten algoritmes en hun mate van transparantie. *Remco Boersma*, projectleider Big Data Living Lab: "Dat gaat van lage complexiteit met een eenvoudige beslisboom (afhandeling van verkeersboetes) tot aan complexere algoritmes (zoals een DNA-verwantschapsonderzoek of gezichts- en beeldherkenning)."

Volgens *Peter van der Putten*, global director Decisioning Solutions bij Pegasystems, denken consumenten ten onrechte dat zij nog nauwelijks te maken hebben met AI: "Spamfilters, virtuele assistenten, aanbevolen nieuwsitems op Facebook: dagelijks gebruiken we al AI-technologie."

Zie ook het verslag op iBestuur.nl: https://ibestuur.nl/nieuws/ai-gaat-over-de-kwaliteit-van-data

op neurale netwerken in het brein. Kunstmatige neuronen in deze netwerken leren continu, en passen hun verbindingen aan op basis van de data die ze verwerken. Spelletjes spelen, plaatjes herkennen, spraak genereren en zelfstandig robots laten lopen: de zelflerende algoritmes kunnen al veel. Een classificatiealgoritme krijgt bijvoorbeeld miljoenen gelabelde foto's van katten en honden te zien. Het leert van deze voorbeelden en zal patronen in de data gaan herkennen die correleren met 'kat' en 'hond', vergelijkbaar met onze ogen en hersenen. Wanneer het op goede data is getraind kan het bij nieuwe foto's, zonder label, met grote zekerheid aangeven welke van de dieren te zien is. Op vergelijkbare wijze kan zo'n zelflerend algoritme worden gebruikt om CT-scans te classificeren als 'kanker' of 'geen kanker'. Het zijn voorbeelden van supervised learning, de meest gebruikte variant, waar de taak van het zelflerend algoritme is om de beste mathematische waardes te vinden die de gelabelde input aan de output verbinden. Omdat gelabelde data vaak duur zijn of ontbreken kan men ook unsupervised learning gebruiken. Hier moet het neurale netwerk zonder interventie van mensen patronen in de data ontdekken. Door het gebrek aan een docent zijn er echter ook geen correcte antwoorden. Ondanks het ongestructureerde karakter zijn deze algoritmes bij uitstek geschikt om onbekende patronen en regels te vinden in grote hoeveelheden bekende data, zoals 'personen die X kopen, kopen ook Y', zonder dat je van tevoren weet waar je eigenlijk naar op zoek bent. AlphaGo Zero, de KI van Google, is op basis hiervan onverslaanbaar geworden in het Chinese bordspel Go. Het systeem ontdekte regels en zetten die mensen nooit gebruikten, maar uitstekend werkten – een soort programmeerbare 'intuïtie'.

Maar welke methode je ook gebruikt, niemand begrijpt wat er nou precies in zo'n model gebeurt. Neem bijvoorbeeld een algoritme dat werd getraind om husky's en wolven te onderscheiden. Op haast alle foto's van wolven in de trainingdata was sneeuw te zien. Uiteindelijk ontdekte het algoritme niet de uiterlijke verschillen tussen de twee dieren, maar de aanwezigheid van sneeuw. Echt begrijpen doen de algoritmes dus nog niet. Waarom classificeerde de Uber de voetganger in luttele seconden als voertuig, fietser en onbekend object? Dat is na afloop moeilijk te zeggen. Het is waarschijnlijk de reden waarom (hopelijk) niemand Google Translate gebruikt om een volledige sollicitatiebrief te vertalen en te versturen. Je snapt niet wat het systeem doet en je kunt de uiteindelijk kwaliteit van je vertaling zonder tolk niet verifiëren. Als gebruiker moet je het maar geloven die goocheltruc, en hopen dat het goed genoeg is.

Een miskwalificatie 'husky-wolf' maakt niet uit. 'Delinquent-brave inwoner' ligt al een stuk gevoeliger. En het helpt niet dat die 'black box' werkt op basis van eerder gepresenteerde data die onvolledig of bevooroordeeld kunnen zijn. Daardoor kan het verleden onterecht het heden gaan bepalen. Dat blijkt uit het functioneren van het COMPAS-algoritme van het bedrijf Equivant. COMPAS zou een snelle, goedkope en objectieve manier bieden om te voorspellen of veroordeelden in de toekomst vaker de fout in zullen gaan. Na analyse van het Darthmouth College bleek de KI echter niet beter te werken dan een groep vrijwilligers. Het systeem discrimineerde bovendien op oneerlijke wijze op etniciteit, waarin het zwarte mannen benadeelde ten opzichte van witte mannen.

Jaap van den Herik, hoogleraar computerwetenschappen en recht aan de universiteit Leiden, heeft onderzoek gedaan naar de validiteit van dit soort digitale rechters. In 2015 demonstreerde hij dat algoritmes in het domein van incassoprocedures, even sterke, of zelfs betere beslissingen dan menselijke rechters namen. Hij beaamt dat er vooroordelen rondzingen, maar ziet desalniettemin een toekomst waarin algoritmes juristen kunnen vervangen. 'Het juridisch systeem is vergelijkbaar met AlphaGo Zero: het vormt zich rond casussen, met menselijke rechters die net als grootmeesters ook fouten hebben gemaakt. Als we alleen de goede intuïtie programmeren, zouden algoritmes op den duur ook betere ethische beslissingen kunnen maken.' Op basis van unsupervised learning superieure intuïtie ontwikkelen klinkt verleidelijk, maar zal nog even op zich moeten laten wachten. Van den Herik voorspelt dat robotrechters die juristen vervangen pas rond 2080 arriveren. "De leermechanismes zijn er nog niet."

Mogelijk schuldig

Maar data, gelabeld of niet, moeten ergens vandaan komen. Wanneer de overheid KI net zoals bedrijven dat doen wil inzetten, zal zij persoonlijke data van inwoners moeten gebruiken. Overheid en inwoner hebben echter een andere relatie dan bedrijf en consument. Dat eerste is vaak geen vrije keuze. Ageren tegen een beslissing van de Belastingdienst zal een stuk moeilijker zijn wanneer de beslissing door een niet-transparant algoritme is genomen. Onderzoekers van de Universiteit Utrecht stellen dan ook in een onderzoeksrapport voor BZK dat verdere ontwikkeling van big data en kunstmatige intelligentie de Nederlandse grondrechten vergaand kan aantasten.

Remco Boersma, werkzaam bij het ministerie van Justitie en Veiligheid, biedt een ontnuchterende blik. "We praten veel over kansen en risico's, maar zodra we doorpraten heeft iedereen het met name over de risico's", vertelt hij. Ongewilde (etnische) profilering, vooroordelen, slechte datakwaliteit, onnavolgbaarheid en weinig transparantie – ze passeren inderdaad vaak de revue. "De kansen en mogelijkheden zijn blijkbaar minder interessant voor discussie", stelt hij, "terwijl daar voor de samenleving veel te halen valt. Neem de zaak Tristan van der Vlist. Iedereen begrijpt dan ineens dat dataanalyse nodig is om dit soort drama's te voorkomen. Vervelend als iemand geen wapenvergunning krijgt terwijl hij hem wel zou mogen krijgen, maar de risico's zijn omgekeerd groter. Bovendien kun je tegen zo'n besluit bezwaar maken." Volledig autonome systemen, zonder menselijke tussenkomst en waar bezwaar maken niet goed mogelijk is omdat de totstandkoming van het besluit niet transparant is, vindt Boersma dan ook een slecht idee: "De uitlegbaarheid en controleerbaarheid mogen nooit uit de formule verdwijnen."

Niet transparante risicoprofilering en big data-analyse kunnen het bijeffect hebben dat inwoners zich altijd 'verdacht' voelen. Een groep belangenbehartigers heeft om die reden de staat gedagvaard om het gebruik van het Systeem Risico-Indicatie (SyRI). Volgens hen wegen hier de risico's zwaarder dan de kansen. SyRi is een algoritme dat zeventien verschillende typen persoonsgegevens gebruikt om te voorspellen waar mogelijk sprake is van belasting-, uitkerings- of toeslagenfraude. In dit geval geen classificatie van hond of kat, maar goede burger of slechte burger. De overheid wil vooralsnog geen inzage geven in de werkwijze of successen van de algoritme, terwijl inwoners wel op een speciale 'mogelijk schuldig'-lijst kunnen belanden zonder dat ze daar weet van hebben. Bezwaar maken wordt daardoor een moeilijk verhaal.

Kunstmatige intelligentie en zelflerende algoritmes: net als elke andere technologie blijken ze een tweesnijdend zwaard. De beslissingen van zelflerende algoritmes zijn in veel domeinen snel, accuraat en objectief, maar kunnen ook grove fouten bevatten, vooroordelen propageren en wantrouwen creëren. Het meer data = beter-principe heeft in zakelijke contexten weliswaar veel successen opgeleverd, voor de overheid gaat het privacytechnisch een lastige zaak worden. Aan ons de burgers om te identificeren waar zelflerende algoritmes en KI onze samenleving écht slimmer en efficiënter maken. Transparantie en inspraak moeten in de verdere ontwikkeling daarom centraal staan, zodat partijen begrijpen wat computer says yes werkelijk betekent. Gelukkig kunnen sommige KI's, zoals Google's DeepMind, al net als mensen uitleggen waarom ze tot een beslissing zijn gekomen. Een wenselijke richting, want als de boel niet transparanter wordt, fietst straks de samenleving met blind vertrouwen en iPhone in de hand de donkerte van de Piet Heintunnel tegemoet.