Publicatie : AG Connect
Datum : 31 jul 2019
cm2 : 1.381
Pagina : 11
Advertentiewaarde : € 16.250,00
Regio : Nederland
Frequentie : 10x per jaar
Oplage : 8.700

De computer zegt nee. Maar waarom?

KIJKEN IN DE ZIEL VAN AI

Kunstmatig intelligente computersystemen vertrouwen we steeds meer taken toe, van medische diagnoses tot energietransport. Maar hoe kun je nog nagaan of deze systemen echt tot de juiste beslissingen komen? Volgens Sander Bohté, Eric Pauwels en Han La Poutré moeten we vóórdat we ze cruciale taken toevertrouwen, doorgronden hoe we AI-systemen zo betrouwbaar en inzichtelijk mogelijk maken.

door Sander Bohté, Eric Pauwels en Han La Poutré illustratie Marc Kolle

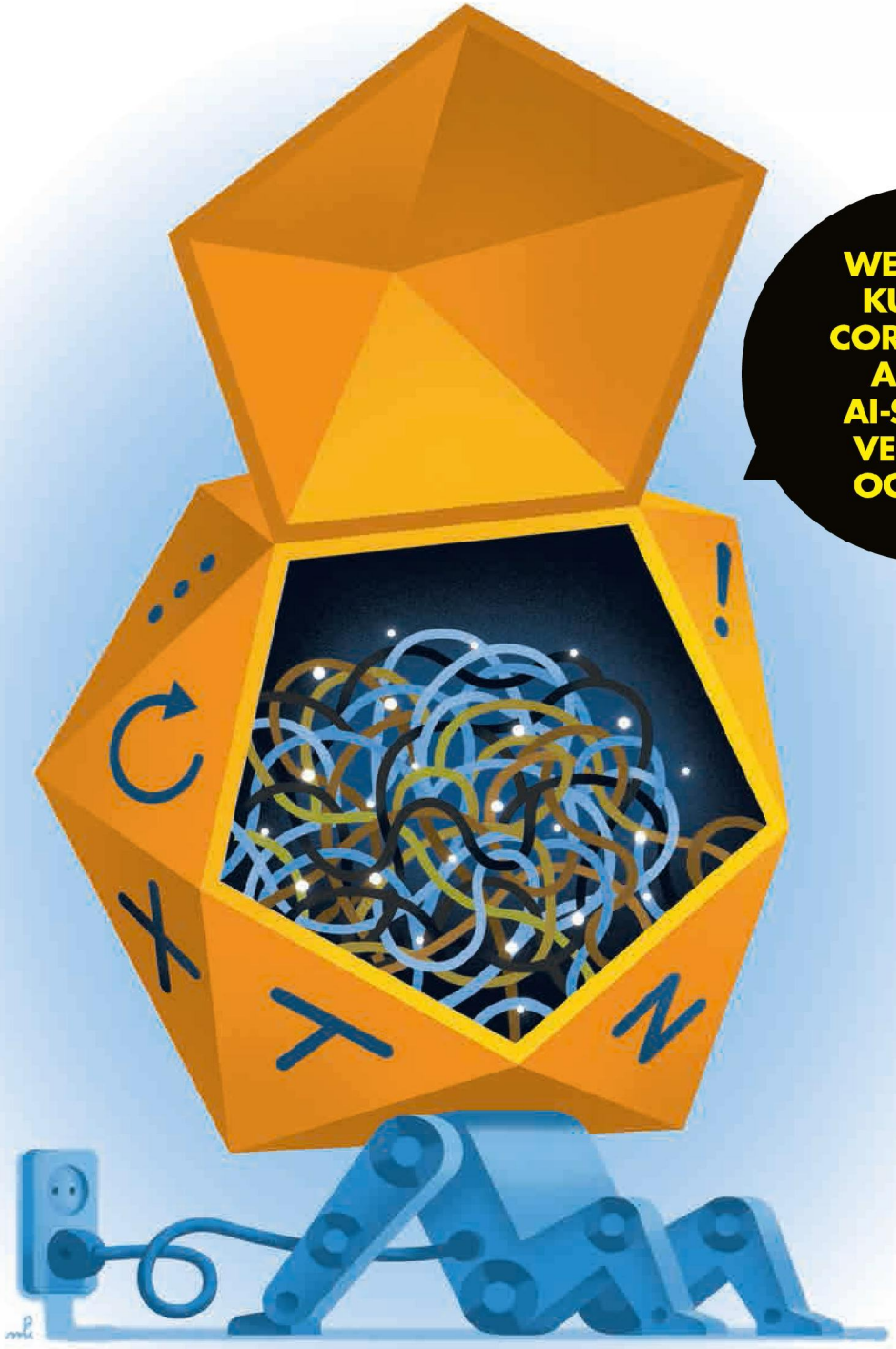
DE KANSEN EN MOGELIJKHEDEN VAN KUNSTMATIGE INTELLIGENTIE GROEIEN MET DE DAG. Langzaam maar zeker krijgen we ook door hoe AI nu eigenlijk werkt, en hoe degelijk het is. Hoe kunnen we AI niet alleen gebruiken, maar ook begrijpen? Is het een vogel, of een vliegtuig, of toch iets heel anders? Een computer die moeiteloos doorheeft wat er op een foto staat en een vogel van een vliegtuig kan onderscheiden, is al lang geen sensatie meer. Als je maar genoeg voorbeelden geeft, kan een kunstmatig intelligent systeem het zelf leren.

AI kan vooralsnog erg slecht zijn in haar extrapolatie

HORRORSCENARIO'S

De vraag is wel hoe AI zich de komende jaren gaat ontwikkelen. Toekomstvisies over AI slaan al snel door in horrorscenario's waarbij computersystemen de dienst uitmaken en de mens niets meer te zeggen heeft. Nu lijkt dat vooruitzicht overdreven, maar mensen kunnen wel terecht bang zijn dat ze geen onderdeel meer zijn van de keten waarin beslissingen worden genomen. Voor een deel ontstaat die angst omdat we willen begrijpen wat er gebeurt en ook willen kunnen corrigeren als een AI-systeem verkeerd oordeelt. Of, in het uiterste geval, ertegen in beroep kunnen gaan. Wat dat betreft staan we op een bijzon-

Publicatie : AG Connect
Datum : 31 jul 2019
cm2 : 1.381
Pagina : 11
Advertentiewaarde : € 16.250,00
Regio : Nederland
Frequentie : 10x per jaar
Oplage : 8.700



Publicatie : AG Connect
Datum : 31 jul 2019
cm2 : 1.381
Pagina : 11
Advertentiewaarde : € 16.250,00
Regio : Nederland
Frequentie : 10x per jaar
Oplage : 8.700

der punt in de ontwikkeling van AI. Terwijl talloze onderzoekers succes boeken door AI-systemen steeds groter en sneller te maken, groeit het besef dat we niet goed weten hoe AI precies werkt en hoe degelijk hun beslissingen zijn. Tegenwoordig is zeer veel onderzoek – en dan vooral bij grote bedrijven – erop gericht om de performance van AI te verbeteren voor allerlei applicaties. Academisch onderzoekers kunnen een belangrijke rol spelen door niet mee te gaan in deze wedloop, maar een stap terug te zetten en vragen wat AI in essentie doet.

ANDERS DENKEN

Een van de moeilijkheden is dat AI-systemen vooralsnog heel anders 'denken' dan mensen. Daardoor kunnen ze niet uitleggen wat ze precies doen – althans niet op een manier die voor een doorsneemens te begrijpen is.
Een AI-systeem 'leert' als het zeer veel voorbeelden krijgt en daarin wetmatigheden kan ontdekken. Het systeem kan dan allerlei interessante generalisaties maken en op basis daarvan beslissingen nemen. Een grote beperking is er wel. AI-systemen zijn vooral goed in het voorspellen van mogelijkheden die vallen binnen de voorbeelden die ze zelf hebben gezien. Verlang je meer van AI-systemen, dan kan het al snel misgaan. AI kan vooralsnog erg slecht zijn in haar extrapolatie, ofwel voorspellen wat er gebeurt buiten het gebied waarin ze voorbeelden hebben gezien.

ZEKERHEID

AI-systemen moeten dus niet enkel krachtiger worden, maar ook inzichtelijker. Een belangrijke stap is helder krijgen hoe betrouwbaar het oordeel van een AI-systeem is. Met hoeveel zekerheid weet je dat dit oordeel correct is? Een van de grootste uitdagingen voor AI-onderzoekers is het opstellen van regels die aangeven hoe groot de fouten marges zijn die een AI-systeem omgeven. Met andere woorden: hoe robuust is het systeem eigenlijk? Wat die uitdaging vooral zo groot maakt, is dat de 'informatieruimte' waarmee een AI-systeem werkt enorm veel dimensies kent. Door die complexiteit is het bijzonder moeilijk om voor alle mogelijke gevallen te bepalen hoe een AI-systeem zal reageren, en hoe groot de foutenmarges zullen (of mogen) zijn. Stel, een medisch AI-systeem heeft geleerd van 100.000 patiëntendossiers en constateert dat jij gezond bent. De volgende dag leert het systeem verder, en krijgt één extra dossier toegevoegd dat afwijkt van alle andere dossiers. Het zou zeer onwenselijk zijn als die ene toevoeging ertoe leidt dat het systeem tot compleet andere oordelen komt – en jou ineens ziek verklaart. In hoeverre kun je dan nog vertrouwen op het oordeel van zo'n AI-systeem?

TAAIE PUZZEL

Het is een typisch probleem van deep learning: ook als je het systeem rare input geeft, krijg je altijd een uitkomst. Maar die uitkomst gaat dan nergens over. AI-onderzoekers willen daar getallen aan kunnen hangen en dus uitdrukken hoe betrouwbaar de uitkomst is. Het is momenteel een van de taaiste puzzels binnen het AI-veld om daar een werkbare standaard in te vinden.
Die puzzel is zo complex omdat het een waar vragenvuur oproept. Als een AI-systeem denkt dat er een vogel op de foto staat, hoe zeker is het daar dan over? Is dat 99%, of slechts 10% of 1%? De vervolgvragen stapelen zich snel op. Zijn bepaalde pixels 99 van de 100 keer hetzelfde? Hoe 'soortgelijk' waren die foto's eigenlijk en hoe meet je dat? Is het de afstand tot het dichtstbijzijnde plaatje waarvan we wel zeker weten dat het een vogel is? Of is het de afstand tot de grens van de categorie 'vogel'?
Hoe meet je die afstand eigenlijk? Kortom, wat zijn goede maten?
Om te begrijpen hoe groot die noodzaak

STROMINGEN IN AI-ONDERZOEK

Onderzoek naar kunstmatige intelligentie (AI), en dan met name machine learning, richt zich niet alleen op het sneller of groter maken van AI-systemen. De afgelopen decennia zijn er stromingen ontstaan binnen het AI-onderzoek die meer inzicht verschaffen in de werking en betrouwbaarheid van AI.

Explainable AI

Een AI-systeem leert en slaat kennis op. Hoe komt het vervolgens tot oordelen of adviezen? Het doel is dat AI-systemen helder kunnen uitleggen wat de structuur o methoden zijn die werden gebruikt om tot een bepaalde beslissing te gekomen.

Accountable AI

Accountable AI is eigenlijk geen goede term. Accountability is een vorm van verantwoordelijkheid. Een persoon of bedrijf is verantwoordelijk en neemt een besluit op basis van de betrouwbaarheid van informatie. Denk bijvoorbeeld aan een zelfrijdende auto met een systeem dat draait op kunstmatige intelligentie. Wanneer kun je veilig met die auto de weg op? Simpel: wanneer die auto beter rijdt dan jij. Maar wie neemt dan de verantwoordelijkheid als er iets misgaat? De centrale vraag bij accountable AI is: wanneer heb je een betrouwbare voorspelling of beslissing en wanneer niet? En kan ik op basis daarvan een beslissing nemen?

Reliable AI

Een ideaal AI-systeem leert snel, maar laat zich niet te snel van de wijs brengen door afwijkende input. Het moet robuust zijn en op een goede manier rekenschap kunnen afleggen over de manier waarop het werkt. Het vakgebied van reliable AI werkt aan manieren om AI betrouwbaar en robuust te maken.

Publicatie : AG Connect
Datum : 31 jul 2019
cm2 : 1.381
Pagina : 11
Advertentiewaarde : € 16.250,00
Regio : Nederland
Frequentie : 10x per jaar
Oplage : 8.700

is, hoef je alleen maar te denken aan de vele grote AI-toepassingen die in aantocht zijn. Een verkeerd beoordeelde vogelfoto heeft misschien geen grote gevolgen. Maar iedereen die een hypotheek aanvraagt, moet wel kunnen begrijpen of die terecht wordt toegekend of afgewezen. Hetzelfde geldt voor een medisch AI-systeem dat een diagnose stelt. Een arts moet kunnen nagaan hoe betrouwbaar die diagnose is. Dat vereist een enorme vertaalslag, omdat het AI-systeem in menselijke termen moet uitleggen wat er is gebeurd, maar in werkelijkheid honderdduizenden 'denkstappen' heeft gezet om tot een uitkomst te komen.

Evenwel is dit de tijd om als IT-sector te investeren in AI

PIXELS

Momenteel zijn er al wel enkele tools beschikbaar die proberen te ontrafelen op basis van welke input het AI-systeem tot een bepaald resultaat is gekomen. Bij bijvoorbeeld beeldclassificatie worden dan de pixels geaccentueerd die het meest hebben bijgedragen aan het classificatiere sultaat. Iets vergelijkbaars kun je doen bij tekstanalyse en de woorden highlighten die aan de basis van het antwoord liggen.

WISKUNDIGE THEORIE

Het probleem met dit soort tools is dat ze ons ook gemakkelijk kunnen misleiden. Menselijke hersenen zijn continu bezig om betekenis te vinden en daarom zien we al snel patronen die er niet zijn. Kijk maar eens hoe gemakkelijk we patronen en vormen zien als we naar de wolken kijken. Als iemand op een plaatje tien willekeurige locaties aanwijst en beweert dat die hebben geholpen om tot een bepaalde classificatie te komen (dit is een vogel), dan zal iedereen wel een paar daarvan als heel zinvol interpreteren. Wat we nodig hebben, is wiskundige theorie die ons vertelt wat de mogelijkheden en limieten van dit soort benaderingen zijn.

DENKEN ALS EEN MENS

Sommige onderzoekers, bijvoorbeeld deep-learninggrondlegger en Turing Award-winnaar Geoffrey Hinton, beweren dat we mogelijk het probleem verkeerd aanpakken. Volgens hen moeten we zoeken naar AI-technologieën die de werking van hersenen beter nabootsen. Menselijke intelligentie denkt meer in termen van concepten. We groeperen dingen en concepten in clusters, en die clusters worden dan de nieuwe 'conceptuele atomen' op een volgend abstractieniveau. De huidige AI-systemen doen dat maar in heel beperkte mate.
Als we AI-systemen kunnen bouwen die dit principe op een structurele manier implementeren, dan staan we weer een stap dichter bij 'explainable AI'. Immers, wanneer wij mensen beweren dat we iets kunnen uitleggen, dan bedoelen we dat we iets kunnen relateren aan andere eenvoudigere concepten. AI-systemen die dit effectief doen, zijn vooralsnog beperkt in kracht.

EXTRA TOOL

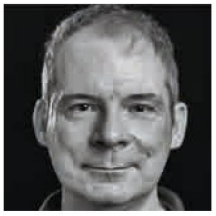
Evenwel is dit de tijd om als IT-sector te investeren in AI, zolang het maar wordt beschouwd als extra tool naast de andere kwantitatieve technieken, zoals statistical learning, optimalisatietechnieken en data-analyse. Met alle kansen en mogelijkheden die het met zich meebrengt, zal AI zich blijven ontwikkelen tot een onmisbare technologie – zeker zolang onderzoekers de tekortkomingen van AI blijven blootleggen en verbeteren. De huidige focus op deep learning, en bijhorende uitdagingen die AI als black box omgeven, groeit dan vanzelf door naar een meer gebalanceerde benadering, waarin we AI gebruiken én de betrouwbaarheid ervan begrijpen.

Dit artikel is ingekort ten behoeve van het magazine. Zie www.agconnect.nl voor de volledige tekst.

AUTEURS

SANDER BOTHE
is senior onderzoeker bij CWI, in de onderzoeksgroep Machine Learning. Daarnaast is hij hoogleraar Cognitive Computational Neuroscience aan de Universiteit van Amsterdam en hoogleraar Biologically Inspired Neural Computation aan de Rijksuniversiteit Groningen.

ERIC PAUWELS
is onderzoeker bij CWI en leidt daar de onderzoeksgroep Intelligent and Autonomous Systems.



HAN LA POUTRE
is onderzoeker bij CWI's onderzoeksgroep Intelligent and Autonomous Systems. Daarnaast is hij hoogleraar Intelligent Energy Systems aan de Technische Universiteit Delft.