



Sophia, de antropeïde robot. © AFP

Laten we zorgen dat kunstmatige intelligentie ons helpt en begrijpt

We moeten Artificiële Intelligentie niet als heilig of duivels zien, maar zorgen dat ze ons helpt en begrijpt.

Adelbert Bronkhorst onderzoekt bij TNO de praktische mogelijkheden van Artificiële Intelligentie (AI) 15 november 2017, 20:53



Blijf op de hoogte

Iedere dag rond lunchtijd het belangrijkste nieuws van de ochtend, de mooiste fotografie en het gesprek van de dag? Schrijf u in voor onze gratis nieuwsbrief.

Go geldt als 's werelds moeilijkste bordspel. Door het enorme aantal mogelijkheden was het winnen van de beste menselijke speler de heilige graal voor Artificiële Intelligentie (AI, kunstmatige intelligentie). De verpletterende zege van 100-0 die AlphaGo Zero, die

zichzelf Go leerde spelen, behaalde op zijn pas één jaar oudere broertje AlphaGo toont definitief aan dat de mens op dit vlak is verdrongen door AI.

Singulariteit

Moeten we ons nu afvragen of de in 2005 door Raymond Kurzweil aangekondigde singulariteit zich niet reeds aan het voltrekken is? En of dat erg is? Singulariteit is de explosieve toename van intelligentie die ontstaat vanaf het moment dat AI zo slim en krachtig is dat ze zichzelf razendsnel kan verbeteren.

Sommigen, zoals Kurzweil zelf, zien het als een transcendentale, in wezen positieve gebeurtenis. Anderen vrezen doembeelden van onderdrukking zoals in *The Terminator* en *The Matrix*. Wie heeft er gelijk? Staan we inderdaad op de drempel van een radicale transformatie van de wereld?

Wanneer je als onderzoeker AI toepast en bestudeert, kun je alleen maar bewondering hebben voor de collega's die systemen zoals AlphaGo ontwikkelen. Anderzijds weet je ook dat AI nog enorm veel 'handwerk' nodig heeft en eigenlijk alleen werkt in gestructureerde en begrensde omgevingen zoals een bordspel.

Er komt nog geen enkel AI-systeem in de buurt van de vaardigheden die 'iemand' als goed teamlid moet hebben: begrip van doelen, behoeften en intenties van anderen.

Singularity lijkt dus niet dichtbij; volgens deskundigen kunnen we het rond 2040 verwachten. Wil dit zeggen dat AI dan pas een radicale transformatie zal veroorzaken? Waar we niet aan twifelen is dat AI al snel de grenzen zal oprekken van wat wij aan machines overlaten. Ons doembeeld is daarbij niet de hyper-intelligente machine die de mens gaat knechten, maar de onvoorspelbare machine die domme, gevaarlijke dingen doet omdat hij in situaties komt die de programmeur niet had voorzien.



Laten we investeren in bruikbare, begrijpelijke en betrouwbare
AI

Vandaar ook de ethische discussie over autonome wapens voor de krijgsmacht, [ook in de Volkskrant wel geframed als 'killer robots'](#). Wapens die adequaat reageren op voorgeprogrammeerde situaties, maar het niet doorhebben wanneer er iets onvoorziens

gebeurt en wat de implicaties van fouten zijn. Waar en hoe blijft de mens in de besluitvorming?

Laten we ons dus niet druk maken over of we AI als heilig of duivels zien, maar investeren in bruikbare, begrijpelijke en betrouwbare AI. AI die zelf kan aangeven wanneer de verstrekte instructies onduidelijk of ambigu zijn. AI die uitlegt waaróm uit de gebruikte modellen en data komt dat op bepaalde plekken extra politiepatrouilles nodig zijn. En natuurlijk ook AI die ons niet per ongeluk tegen een vrachtwagen of huis laat crashen.

Dat betekent dat we als onderzoekers en ontwikkelaars van AI niet alleen slimme dingen moeten bedenken, maar ook goed moeten begrijpen hoe de AI wordt toegepast en welke fouten daarbij kunnen optreden.

Kaderen

Dat is niet zo spannend als werken aan superintelligente bordspelcomputers, maar even degelijk als eerder werk om andere riskante technologie, zoals vliegtuigen, bruikbaar en veilig te maken.

Hierbij moet worden voorkomen dat AI sterk wordt ingekaderd. Dan verkleinen we ook de kans dat machines ons gaan verrassen met prachtige ideeën en oplossingen waar we zelf niet op hadden kunnen komen.

Als we dan per se een utopisch toekomstbeeld willen realiseren, laten we dan ook mensvriendelijke human-aware AI nastreven, die ons als een goede vriend of collega begrijpt, helpt en ondersteunt. AI die ook sociale intelligentie heeft en dus heel goed snapt dat je het niet leuk vindt als hij je in een spelletje met 100-0 vernedert - en dan het risico loopt dat je er gewoon de stekker uittrekt.



Volg en lees meer over:

★ **ROBOTS** | ★ **TECH** | ★ **OPINIE**

ROBOTANGST

67 ARTIKELEN
