

Computer kan beter liplezen dan professional

Laurens Verhagen
 Amsterdam

Opnieuw heeft de computer een overwinning geboekt op de mens. Engelse wetenschappers hebben samen met Google een computer gemaakt die beter kan liplezen dan professionals.

Was het eerder dit jaar de wereldkampioen Go die zijn meerdere moest erkennen in een zelflerende computer van Google, dit keer zijn het de professionele liplezers die eraan moeten geloven. Onderzoekers van de Universiteit van Oxford zijn er samen met Googles kunstmatige intelligentieprogramma DeepMind in geslaagd computers behoorlijk accuraat te laten liplezen.

'Behoorlijk accuraat' is een slagingspercentage van 46 procent van alle woorden bij video's van tv-programma's die DeepMind niet eerder had gezien. Dit klinkt voor de leek wellicht weinig overtuigend, maar ook minieme foutjes als een ontbrekende 's' worden de liplezer zwaar aangerekend. Met 46 procent doet Google het stukken beter dan ingehuurde professionals, die niet verder komen dan 12,4 procent van de woorden.

Sander Bohte doet aan het Centrum Wiskunde & Informatica in Amsterdam onderzoek naar diepe neurale netwerken en *machine learning* en is onder de indruk van dit resultaat. 'Dat de machine zoveel beter presteert dan de mens, is echt spectaculair.'

Het is volgens Bohte niet verrassend dat deze nieuwe doorbraak ook weer met DeepMind gebeurt. Sinds Google die in 2014 overnam, is het hard gegaan. Nu werken in Londen honderden kunstmatige intelligentie-experts. In combinatie met de gigantische computerfaciliteiten van Google leidt dit tot veel opvallende resultaten. Bohte: 'Google is natuurlijk niet de enige die hiermee bezig is, maar Google DeepMind is op dit moment het verst met kunstmatige intelligentie.'

Zoals bij alle kunstmatige intelligentie experimenten van Google geldt het adagium 'oefening baart kunst'. Dit is

dan ook een van de redenen van Googles bijna niet te stillen datahonger. In dit geval bood de BBC uitkomst. De onderzoekers voedden hun computer met vele duizenden uren video van bestaande Britse tv-programma's met pratende hoofden.

De BBC-programma's waren goed voor ruim 118 duizend verschillende zinnen en 17.500 unieke woorden, afkomstig van zo'n duizend sprekers. Ter vergelijking: een ander onderzoeksprogramma in Oxford, genaamd LipNet, moet het doen met duizend zinnen van 34 verschillende sprekers. Dit LipNet gebruikte ook nog eens videomateriaal dat speciaal voor dit doelinde was opgenomen. Geen presentatoren die door elkaar praten of woorden inslikken, maar mensen die duidelijk articulerende zinnen inspreken.

Googles Deep Mind werd gevoed met 118 duizend zinnen

DeepMind doet zijn werk in een veel natuurlijker omgeving, wat de weg vrijmaakt voor allerlei praktische toepassingen. De onderzoekers noemen in hun artikel de mogelijkheid van het geven van instructies in een mobiele telefoon in een lawaaiige omgeving. Denk aan Siri, maar dan beter. 'Sorry, ik denk niet dat ik u helemaal verstond' is er dan niet meer bij. Een andere toepassing is directe transcriptie van uitgesproken teksten, bijvoorbeeld voor ondertitels. 'Als je dat combineert met spraakherkenning, kom je nog verder dan nu al het geval is', stelt Bohte.

Her en der is al geopperd dat geheime diensten ook wel veel belangstelling zullen hebben voor de liplezende computer. Een vertrouwelijk gesprek in een mensenmassa is er dan niet meer bij. Of is dat sciencefiction? 'Nee hoor', zegt Bohte. 'Dat is prima mogelijk als de kwaliteit van de beelden maar hoog genoeg is.'

