

# Kunstmatige intelligentie kan vooroordelen hebben?

15 april 2017

Tijdens de studie maakte de onderzoekers gebruik van de Implicit Association Test (IAT), een cognitieve reactietaak die ontworpen is om de sterkte van associaties tussen begrippen in het geheugen te onderzoeken. De test wordt vaak toegepast om onbewuste vooroordelen, of impliciete voorkeur ten aanzien van groeperingen in de maatschappij te meten. Deze test wordt sinds eind jaren '90 in veel psychologische onderzoeken gebruikt.

De onderzoekers gebruikten voor de studie een programma dat functioneert als de machine learning-versie van de IAT. Dit programma genaamd GloVe is ontwikkeld door wetenschappers aan Stanford University en is gebaseerd op algorithmen die zogeheten word vectors kunnen bouwen.

Word vectors binnen machine learning zijn grote verzamelingen van nummers die het concept van een bepaald woord kunnen weergeven. Wetenschappers denken dat deze methode van fundamenteel belang zijn in hoe computers in de toekomst grote hoeveelheden tekst kunnen analyseren en interpreteren.

## **'Machine learning leren begrijpen wat wel en niet acceptabel is'**

Volgens de onderzoekers zijn AI-systemen allesbehalve objectief of onbevooroordeeld alleen maar vanwege het feit dat zij gebaseerd zijn op wiskunde en logica. Aan het begrip 'insecten' worden bijvoorbeeld vaak negatieve kenmerken en aan het begrip 'bloemen' positieve kenmerken toegekend. Op dezelfde manier kunnen positieve of negatieve eigenschappen worden geassocieerd met bepaalde maatschappelijke groepen, zoals personen met een verschillende etnische achtergrond, of personen van verschillende leeftijd.

"De vooroordelen en stereotypen in onze maatschappij komen tot uiting via onze spraak. Deze systemen zijn complex en bovendien hardnekkig. Het is dan ook weinig zinvol om deze systemen te saneren of te elimineren. In plaats daarvan moet een methode worden ontwikkeld om machine learning te leren begrijpen wat wel en niet acceptabel is", aldus de onderzoekers.