Onnodige angst voor algoritmen

‘Onderzoekers bij Facebook waren geschokt toen een AI-experiment uit de hand dreigde te lopen. De wetenschappers voelden zich genoodzaakt de stekker uit twee robots te trekken nadat deze stiekem een geheime taal hadden ontwikkeld.’ De kans is groot dat u dit verhaal vorig jaar voorbij hebt zien komen. Een soortgelijk bericht werd massaal gedeeld en deed even later ook de ronde in een versie waarbij Facebook was vervangen door Google.

Het verhaal is echter lariekoek. Wie teruggaat naar de bron kan lezen dat de onderzoekers helemaal niet zo ongerust waren. Het ging om een doodnormale chatbot (zoals Facebook er zoveel heeft) die ze probeerden te laten onderhandelen. De bot moest zelf bepalen hoeveel hoeden, ballen of boeken hij bereid was te ruilen in een chatconversatie. Om hem sneller te laten leren, lieten ze de bot oefenen met een andere chatbot. Wat de onderzoekers hadden nagelaten, was de bots zodanig te programmeren dat ze in normale volzinnen zouden praten.

De onderhandeling ging als volgt. Alice: ‘Ik wil boek-boek-boek en hoed-hoed’. Bob: ‘Dat is goed, maar dan wil ik bal-bal-bal-bal-bal.’ Alice: ‘Deal!’ Lekker efficiënt, toch? Maar deze manier van converseren was voor de onderzoekers niet bruikbaar. Het was immers de bedoeling dat Alice en Bob ook met echte mensen zouden kunnen praten. Dus —zo concludeerden de onderzoekers —was dit experiment mislukt. Laten we het op een andere manier proberen.

Ik heb nog nooit berichten gelezen zoals: calculator rekent ’s nachts stiekem eigen sommen uit

Dat is echter niet wat de journalisten en twitteraars in dit bericht zagen. De doemscenario’s werden weer opgepoetst en rijkelijk geïllustreerd met foto’s van de Terminator het wereldwijde web op geslingerd. De rectificatie waarin de wetenschappers een genuanceerd verhaal vertelden en zelfs het onderzoek volledig publiceerden, kreeg nauwelijks aandacht.

Een dergelijke trend doet zich nu ook voor rond algoritmen. Er wordt gewaarschuwd dat algoritmen de baas over ons willen spelen. En dat is slecht, want ze zijn ‘oneerlijk’. Weer lariekoek! Algoritmen hebben geen eigen wil. Ze zijn al jaren onderdeel van het dagelijks leven. Zonder algoritmen zou TomTom de weg niet kunnen vinden naar uw huis. Stoplichten zouden niet werken. Uw rekenmachine zou niet eens 1 + 1 kunnen uitrekenen. Toch heb ik nog nooit nieuwsberichten gelezen als: calculator rekent ’s nachts stiekem eigen sommen uit. Of: verkeerslicht zet bewust alle lichten op groen.

We moeten oppassen met antropomorfisme. Door algoritmen menselijke eigenschappen toe te dichten, lijkt het of een deel van de verantwoordelijkheid ook naar de machines verschuift. Algoritmen zijn niet de vijand. Ze zijn geavanceerd gereedschap in handen van mensen zoals u en ik. Dus ja, er schuilt gevaar in geautomatiseerde besluiten. We moeten voorkomen dat menselijke waarden geplet worden tussen de digitale walsen. Maar de ‘bias’ die we algoritmen kwalijk nemen is niets anders dan een optelsom van onze eigen historische vooroordelen en voorkeuren. Het is als boos worden op een spiegel omdat je haar niet goed zit.

De uitdaging van de komende jaren is nu om systemen te ontwerpen die onze eigen onvolmaaktheden neutraliseren. Wie zich daarvoor inzet, krijgt van Alice en Bob een dikke like. Deal!