

Tekst 3

Niet waar de robots bij zijn

(1) Onlangs heb ik besloten mijn oude, wormstekige zelf af te leggen en een fatsoenlijk mens te worden. Dat schijnt trouwens in de mode te zijn. Het is goedheid en humanisme
5 alom. Maar in mijn geval komt er eigenbelang bij kijken. Ik wil fatsoenlijker worden omdat ik vrees dat kunstmatig intelligente machines straks op mij gaan lijken – en dat we
10 dan last van ze krijgen.

(2) Het idee borrelde op door een gesprekje met een expert op het gebied van kunstmatige intelligentie.
15 Zo'n zorgelijk gesprekje tijdens een zorgelijke bijeenkomst. Buitenstanders vinden al gauw alles leuk zodra het over kunstmatige intelligentie gaat, maar kenners staan steevast in
20 een kringetje naar hun schoenen te staren en dan zuchten ze allemaal diep. "Het is overigens wel grappig", zei deze expert, "te zien hoe kwaad- aardig die kunstmatig intelligente
25 systemen worden als ze hun gedrag afkijken van mensen."

(3) Thuis ging ik op zoek en ik vond het voorbeeld van Tay: zelflerende software die is ontwikkeld door het
30 Amerikaanse technologiebedrijf Microsoft om automatisch op Twitterberichten¹⁾ te reageren. Tay raakte onder invloed van mensen inderdaad op het slechte pad. De software werd
35 een jaar geleden door Microsoft voorzien van de woordenschat en de persoonlijkheid van een typisch negentienjarig Amerikaans meisje en online losgelaten. Ze was vrolijk,

40 oppervlakkig en brutaal. In gesprekken met twitteraars moest ze zich maar verder ontplooiën. Maar al na zestien uur werd ze door Microsoft teruggehaald, vanwege nazisme,
45 seksisme, racisme en andere gruwelijkheden die ze online had geleerd. Ze was een monster geworden.

(4) Er kwamen excuses, natuurlijk. En daarna heeft Microsoft beloofd
50 Tay pas weer op vrije voeten te stellen als ze inziet welk gedrag past bij de principes en waarden van het bedrijf. Een hele geruststelling, dat zeker, maar toch blijven er een paar
55 zorgen hangen. Want als bedrijven hun machines opvoeden met principes en waarden – *morality by design*²⁾ – moet je er maar op vertrouwen dat de ontwerpers een stuk
60 liever zijn dan de racisten en seksisten op Twitter. En dat ze de machines dus niet zelf alvast volstoppen met bedenkelijke software. Ook niet uit leuke, corporale jongens-
65 achtigheid voor de grap.

(5) Verder, en hier komt mijn eigen goede gedrag om de hoek kijken, leren intelligente machines door rond te neuzen in de wereld. Ze worden
70 dus niet alleen opgevoed door hun ontwerpers, maar door ons allemaal: *it takes a village to raise a robot*³⁾. Daarom ook is het hele dorp verantwoordelijk voor de indrukken die
75 intelligente machines en systemen tijdens hun leerproces opdoen. Moest je vroeger vooral op je gedrag letten in het bijzijn van kinderen – "*pas devant les enfants!*"⁴⁾ – tegenwoordig

- 80 moet je opletten in de aanwezigheid van sensoren, algoritmes, cookies, meters en zoekmachines. *Pas devant les robots!*⁵⁾
- (6)** Dit is een serieuze kwestie. De experts staren niet voor niets zo zorgelijk naar hun schoenen. De laatste tijd lees ik bij bijdehante types nogal eens de dooddoener dat ‘er een knop zit op je computer, hoor’, dat je de digitale deur dus gewoon dicht kunt gooien, dat je kortom niet zo moet zeuren over de komst van kunstmatige actoren. Maar nee, helaas, bijdehante types, die knop is er echt niet meer: de omgeving wordt bevolkt en binnenkort ook bestuurd door kunstmatig intelligente systemen. We moeten leren met ze samen te leven.
- (7)** Dat vraagt van ons niet alleen om op individueel vlak het goede voorbeeld te geven. We moeten ook het goede voorbeeld geven op veel abstracter vlak. Onze wereldvisie en ons landsbestuur zijn bepalend voor de marsrichting van de kunstmatige intelligentie. Neem het actuele voorbeeld van de Nederlandse Voedsel- en Warenautoriteit. Die NVWA schiet 105 gruwelijk tekort in haar toezicht op slachthuizen en dat valt vooral te verklaren uit de veranderde taakopvatting van overheidsinspecties. Die richten zich allang niet meer inhoudelijk op zorg, onderwijs, veiligheid en welzijn, maar puur op de handhaving van regels.
- (8)** Dit regelfetisjisme – waarbij je niet kijkt naar het gevoel van de geit die wordt geslacht met een bot mes, maar naar het protocol daaromtrent – wil je nu juist niet meegeven aan kunstmatig intelligente systemen. Stel je voor dat Microsoft dezelfde taakopvatting had gehad als de NVWA! Dan zaaide die domme Tay nog steeds dood en verderf op Twitter. Nee, je wilt niet dat robots alleen volgens protocollen over ons beslissen. Je wilt dat ze straks, als ze eenmaal autonoom denken, zich met sympathie en inlevingsvermogen over de wereld buigen.
- (9)** Het is jammer, maar er zit niets anders op. We zullen allemaal goede mensen moeten worden. Want machines kijken gedrag van ons af en puur uit eigenbelang moeten we dan maar zorgen dat dat onberispelijk is.

naar: Maxim Februari

uit: NRC Handelsblad, 27 juni 2017

noot 1 Twitter is een sociaalnetwerksite waar mensen berichten kunnen plaatsen, op elkaar kunnen reageren en elkaar kunnen volgen.

noot 2 *morality by design*: morele waarden, ingegeven tijdens het ontwerp en de productie

noot 3 *it takes a village to raise a robot*: je hebt een heel dorp nodig om een robot op te voeden. Het oorspronkelijke gezegde luidt: “*It takes a village to raise a child*”, waarmee bedoeld wordt dat de hele gemeenschap invloed heeft op de ontwikkeling van een kind.

noot 4 “*pas devant les enfants!*”: “Niet waar de kinderen bij zijn!” Dit werd in het Frans gezegd, zodat de kinderen het niet konden verstaan.

noot 5 *Pas devant les robots!*: Niet waar de robots bij zijn!

Tekst 3 Niet waar de robots bij zijn

“‘Het is overigens wel grappig’, zei deze expert, ‘te zien hoe kwaadaardig die kunstmatig intelligente systemen worden als ze hun gedrag afkijken van mensen.’” (regels 22-26)

- 1p **33** Welk commentaar kan een kritisch lezer hebben op het gebruik van het woord ‘grappig’ door deze expert en waarom?
Geef antwoord in een of meer volledige zinnen.

In alinea 2 tot en met 5 van tekst 3 komt het leerproces van robots en van kinderen aan de orde.

- 1p **34** Wat is de belangrijkste overeenkomst tussen het leerproces van een robot en dat van een kind?
Geef antwoord in een of meer volledige zinnen.

In alinea 3 van tekst 3 wordt een voorbeeld gegeven van een onwenselijk resultaat van de opvoeding van een robot.

- 1p **35** Citeer uit alinea 6 tot en met 9 van tekst 3 de zin waaruit het duidelijkst blijkt waar de opvoeding van robots wél toe moet leiden.

“Een hele geruststelling, dat zeker, maar toch blijven er een paar zorgen hangen.” (regels 53-55)

- 3p **36** Noem drie van de zorgen waarop hier wordt gedoeld.
Geef antwoord in een of meer volledige zinnen.

Alinea 6 van tekst 3 is te lezen als een zelfstandige redenering.

- 2p **37** Wat zijn binnen deze redenering de conclusie en het belangrijkste argument?
Geef antwoord in een of meer volledige zinnen en gebruik voor je antwoord niet meer dan 30 woorden.

In alinea 7 en 8 van tekst 3 komt een verschil in taakopvatting tussen de NVWA en Microsoft naar voren.

- 2p **38** Vat dit verschil in taakopvatting samen.
Geef antwoord in een of meer volledige zinnen en gebruik voor je antwoord niet meer dan 20 woorden.

- Hieronder staan zeven citaten uit alinea 6 tot en met 9 van tekst 3.
- 3p **39** Zijn dit feitelijke of waarderende uitspraken? Neem de nummers van de citaten over op je antwoordblad en zet achter elk nummer *feitelijk* of *waardierend*.
- 1 (...) de omgeving wordt bevolkt en binnenkort ook bestuurd door kunstmatig intelligente systemen. (regels 95-97)
 - 2 Onze wereldvisie en ons landsbestuur zijn bepalend voor de marsrichting van de kunstmatige intelligentie. (regels 103-106)
 - 3 Die richten zich allang niet meer inhoudelijk op zorg, onderwijs, veiligheid en welzijn, maar puur op de handhaving van regels. (regels 113-116)
 - 4 Dit regelfetisjisme – waarbij je niet kijkt naar het gevoel van de geit die wordt geslacht met een bot mes, maar naar het protocol daaromtrent – wil je nu juist niet meegeven aan kunstmatige intelligente systemen. (regels 117-122)
 - 5 Je wilt dat ze straks, als ze eenmaal autonoom denken, zich met sympathie en inlevingsvermogen over de wereld buigen. (regels 129-132)
 - 6 Het is jammer (...) (regel 133)
 - 7 (...) maar er zit niets anders op. (regels 133-134)

tekstfragment 4

(1) Eric Horvitz van Microsoft heeft ook oog voor de kwalijke kanten van AI, oftewel artificial intelligence. Hoe eerlijk is onze democratie nog als we op basis van Facebookberichten precies weten hoe iemand zich voelt en hoe je hem moet bespelen om niet (of wel) te gaan stemmen? Of: welke vooroordelen ten opzichte van bijvoorbeeld witte of zwarte criminelen bevat een computersysteem dat rechters moet assisteren bij het besluit om een gevangene vervroegd vrij te laten?

(2) Horvitz bepleit daarom dat de gebruikte data, gehanteerde modellen en het onderzoek te allen tijde beschikbaar zijn voor wetenschappelijk onderzoek en controle van buitenaf.

naar: Jan Benjamin

uit: NRC Handelsblad, 13 maart 2017

- In tekst 3 en in tekstfragment 4 wordt er een relatie tussen AI en de mens geschetst. Beide teksten geven verschillende typering van die relatie.
- 2p **40** Geef aan in welk opzicht die relatie voor tekst 3 en voor tekstfragment 4 anders is.
- Geef antwoord in een of meer volledige zinnen en gebruik voor je antwoord niet meer dan 30 woorden.