

Aristoteles in tijden van AI

ALDUS

Dr.ir. Marc Steen werkt als senior onderzoeker bij TNO, onder meer in het programma Verantwoorde Waardecreatie met Big Data.

In Europa willen we kunstmatige intelligentie niet gebruiken om spullen te verkopen of burgers in de gaten te houden, maar om maatschappelijke problemen op te lossen. Daarbij moeten we onze toevlucht nemen tot de deugdeethiek, betoogt TNO-onderzoeker Marc Steen.

Je komt *machine learning*, algoritmes en artificiële intelligentie (AI) steeds vaker tegen, in apparaten, diensten en processen. Dat roept allerlei maatschappelijke en ethische vragen op: over de verslavende werking van social media, over de invloed van bots op verkiezingen, over de winsten van techbedrijven, over surveillance. In Europa willen we daar anders mee omgaan dan in de VS, waar bedrijven AI inzetten om spullen te verkopen, en anders dan in China, waar de overheid AI inzet om burgers in de gaten te houden. Hier willen we AI gebruiken om maatschappelijke problemen op te lossen. En daarbij willen we waarden zoals vrijheid, gelijkwaardigheid, solidariteit, privacy en democratie overeind houden.

Om AI in goede banen te leiden, pleit ik voor de inzet van deugdeethiek. Daarin volg ik Shannon Vallor, hoogleraar aan de Santa Clara University in Silicon Valley. In haar boek *Technology and the virtues* betoogt ze dat we deugdeethiek nodig hebben bij vragen over opkomende technologieën zoals AI.

Deugdeethiek is een alternatief voor plichtethiek en gevolgenethiek. Bij plichtethiek ga je op zoek naar plichten en regels die voor iedereen, altijd en overal gelden. Bij gevolgenethiek probeer je positieve gevolgen te maximaliseren en negatieve gevolgen te minimaliseren. Maar doordat AI volop in ontwikkeling is, zijn die plichten en regels nog niet helder. Bovendien is het lastig om de reikwijdte van de gevolgen in te schatten. Een

zelfrijdende auto heeft niet alleen gevolgen voor de bestuurder en inzittenden, maar ook voor voetgangers en fietsers, en voor je gezin en je sociale leven. Je gaat bijvoorbeeld op het platteland wonen en vier uur heen en vier uur terug forensen naar je kantoor in Brussel. Je kunt immers slapen achter het stuur.

Wat willen we voor maatschappij? Hoe ziet het goede leven eruit? Dat soort vragen stel je vanuit de deugdeethiek. Aristoteles, een van de grondleggers, schreef daar het volgende over. Deugdeethiek veronderstelt dat er doelen bestaan. Een beukenootje heeft als doel om uit te groeien tot een beukenboom. Wij mensen hebben als doel om onze vermogens te ontwikkelen, zoals reflecteren, communiceren, samenwerken en creëren, zodat we betekenisvol kunnen leven. Daarom gaat deugdeethiek over het inrichten van een maatschappij waarbinnen mensen hun vermogens kunnen ontwikkelen.

Onrechtvaardige situaties

Verder gaat deugdeethiek over het cultiveren van deugden en over het vinden van het juiste midden in iedere situatie. Stel dat ik een vechtpartij zie op straat. Als ik niet goed ter been ben, handel ik moedig als ik aan de kant blijf staan en 112 bel. Ingrijpen zou overmoedig zijn. Maar als ik sterk ben en een vechtsport beoefen, handel ik moedig als ik ingrijp. Niets doen zou laf zijn.

Nu hebben wij andere deugden nodig dan de mensen in Aristoteles' Athene. Daarom beschrijft

Vallor twaalf techno-morele deugden die voor ons van belang zijn. Vier voorbeelden.

Ten eerste zelfbeheersing. Als je werkt aan een algoritme, kan de verleiding groot zijn om data uit allerlei bronnen te halen, ze aan elkaar te koppelen en daar allerlei analyses op los te laten. Zo'n gulzigheid kan echter uitdraaien op een puinhoop en op illegale praktijken. Als je zelfbeheersing wilt cultiveren, streef je naar minimaal gebruik van data en naar algoritmes die je kunt begrijpen en uitleggen, zoals 'als A, dan X'.

Je moet ook de moed hebben om soms de stekker eruit te trekken

Ten tweede moed. Deze deugd gaat over het vinden van het juiste midden tussen te veel vertrouwen in technologie en te veel angst voor technologie. Je kunt dat bijvoorbeeld doen door een algoritme zorgvuldig te beproeven, in kleine stappen, in experimenten, en alles kritisch te monitoren en bij te sturen. Daarbij moet je dan ook de moed hebben om de stekker eruit te trekken als de voordelen niet opwegen tegen de nadelen.

Ten derde rechtvaardigheid. Het boek *Weapons of math destruction* van Cathy O'Neil staat vol voorbeelden van hoe algoritmes allerlei onrechtvaardige situaties in stand houden of zelfs verergeren. Je kunt rechtvaardigheid cultiveren door *bias* in trainingsdata te vermijden. Dat deed Google niet toen het bedrijf in 2015 een foto van twee don-

kergekleurde tieners labelde als gorilla's. Het verantwoordelijke algoritme was getraind op foto's van lichtgekleurde mensen.

Ten vierde bescheidenheid. Stel dat je werkt aan een algoritme dat fraudeurs opspoor. Dat algoritme produceert ook foute voorspellingen: soms wijst het iemand aan als fraudeur die geen fraudeur is (*false positive*), soms wijst het iemand aan als niet-fraudeur die wel fraudeert (*false negative*). Je kunt bescheidenheid cultiveren door een proces in te richten waarmee mensen dit soort fouten kunnen opsporen en corrigeren. Of je kunt de onzekerheid in beeld brengen, bijvoorbeeld door de mate van (on)zekerheid bij een uitkomst af te beelden.

Hoe kun je nu aan de slag met deugdeethiek? Je kunt starten met de vraag: welke deugden heb ik nodig in dit project? Daarna ga je die deugden oefenen. Je kunt 'nee' zeggen tegen een feature (zelfbeheersing). Of aan de bel trekken als je ziet dat een algoritme onrechtvaardige resultaten produceert (moed). Ook kun je ervaringen uitwisselen. En je kunt leren van mensen die een of meer deugden belichamen in hun werk; op de site DearEngineer.eu staan voorbeelden. Deugdeethiek is bij uitstek een ethiek voor ingenieurs; zij willen meehelpen bouwen aan een rechtvaardige samenleving waarin mensen kunnen floreren. |