

START

MET VOORUITKIJKEN

Whitepaper

Dilemma's rondom zelfdenkende technologie

Een modern Trolleyprobleem: Wie mag er blijven leven?



Noordhoff

Brengt je verder

Over de auteur

Mr. dr. Bart Wernaart (1983) is docent recht en ethiek aan Fontys Hogescholen, International Business. Daarnaast is hij als docent-onderzoeker verbonden aan het Fontys Lectoraat Smart Society and Strategy, en doet hij postdoc-onderzoek binnen de onderzoeksbeweging Smart Society. Hij publiceert in binnen- en buitenland over vraagstukken op het snijvlak van internationaal recht en ethiek en is bij Noordhoff auteur van de boeken [Ethiek en economie, een grensoverschrijdende inleiding](#); [International law and business, a global introduction](#) en zijn meest recente werk [Ethics and business, a global introduction](#).



Het debat rondom zelfdenkende technologie

De discussie over zelfdenkende technologie en de impact die dit heeft op ons dagelijkse leven, kent veel gezichten. Denk aan de algoritmen die toegepast worden in social media, de in ontwikkeling zijnde chat-bot die we aan de lijn krijgen bij een bouwmarkt of de politie, de seks-bot, de war-bot of de zelfrijdende auto. Deze technologische ontwikkelingen brengen nieuwe, ethische vraagstukken met zich mee voor mensen, bedrijven en overheden.

Zo wordt in het actuele debat rondom de zelfrijdende auto gevraagd wat er moet gebeuren wanneer de auto in een noodsituatie een ethische keuze moet maken (Nyholm & Smids, 2016). Stel je voor: de zelfrijdende auto bevindt zich in een situatie waarin die hoe dan ook een crash gaat veroorzaken. Ze kan daarbij slechts drie richtingen op sturen. In de weg naar links zal de auto een fietser zonder helm, en in de weg naar rechts een fietser met helm aanrijden. De weg rechtdoor zal alle inzittenden van de auto het leven kosten, omdat hier een betonblok staat. Er is geen andere oplossing mogelijk. Wat moet de auto beslissen? En misschien nog wel belangrijker: wie beslist over de exacte ethische programmering van de auto waaruit de beslissing voortvloeit?

Morele autoriteit

Om met die laatste vraag te beginnen: er is een complex netwerk aan belanghebbenden dat hier wat van vindt. Zo is er op de eerste plek het bedrijf dat de auto's wil verkopen en de investering in de vernieuwende technologie met winst wil terugverdienen.

Daarnaast is er de programmeur, of een team van programmeurs, die met gewenste de programmering een technisch hoogstandje moeten leveren en daar ongetwijfeld zelf een ethisch gevoel bij hebben. We hebben te maken met een consument, die liever geen auto wil kopen die de gebruiker mogelijkerwijs opoffert, en de samenleving in zijn algemeenheid, die een zo veilig mogelijk verkeer nastreeft.

De vraag is: wie van deze belanghebbenden heeft de morele autoriteit om hier een definitief oordeel over te hebben?

In theorie zouden we dit vrij eenvoudig kunnen oplossen (Millar, 2017). We zouden ten eerste een onderscheid kunnen maken tussen 'moral hard cases' en feitelijke zaken. Iets is immers alleen een moreel dilemma wanneer je het niet met feitelijke zaken kunt oplossen. Een auto dusdanig programmeren dat die zich aan de maximumsnelheid houdt, de snelste of de kortste weg kiest, is een feitelijke programmering waar geen morele dilemma's uit voortvloeien. Vanuit ethisch opzicht is het geen enkel probleem dit over te laten aan de programmeur, in dienst van het producerende bedrijf. De situatie waarbij de auto moet kiezen tussen drie botsingen waarbij personen om het leven kunnen komen is veel minder feitelijk van aard: ook al ben je op de hoogte van alle feitelijke zaken in deze situatie, dan nog is het een kwestie van het afwegen van ethische waarden teneinde een keuze te kunnen maken.

- Kies je ervoor degene met de beste bescherming aan te rijden, omdat dit leidt tot de grootste overlevingskans van alle betrokkenen?
- Kies je ervoor degene aan te rijden die niet diens verantwoordelijkheid nam om veilig deel te nemen aan het verkeer door geen helm te dragen?
- Of offer je de inzittenden van de auto op, aangezien het ten slotte hun autoriteit is met de door hun gekozen bestemming in het door hun gekozen voertuig?

De verantwoordelijkheid van het bedrijf

Bovengenoemde casus is een duidelijk voorbeeld van een 'moral hard case', waarbij de gevolgen 'high stake' zijn: het gaat letterlijk over leven en dood. In dat geval zou het redelijk zijn degene wie het betreft zeggenschap te geven in de morele programmering van de zelfrijdende auto. Aangezien deze beslissingen de verkeersveiligheid in zijn algemeenheid betreffen, en de impact dus maatschappelijk van aard is, lijkt de wetgever de aangewezen persoon hier wat van te vinden. Mochten de beslissingen nu vooral individueel van aard zijn – denk daarbij aan de vraag of de inzittende gewaarschuwd wil worden vlak voor een ongeluk, of liever in het ongewisse blijft – dan ligt het voor de hand de gebruiker te laten kiezen. Dat zou kunnen door de gebruiker de fabrieksinstellingen te laten personaliseren, net zoals we dat soms bij een smartphone kunnen doen.

Er zijn ook 'low stake' moral hard cases. Denk daarbij aan de vraag of de auto bij het kiezen van de snelste route ook de CO₂-uitstoot mee moet wegen, waarbij het verschil in CO₂-uitstoot tussen de mogelijke opties erg klein is. Vanuit een praktisch oogpunt zouden we kunnen betogen dat de programmeur hierin de vrije hand kan krijgen: de impact is erg klein, maar wel van ethisch belang, en kan met een gerust hart aan de 4 ontwikkelaars worden overgelaten, al dan niet in opdracht van het producerende bedrijf.

Echter, de vraag is in hoeverre deze toedeling van ethische besluitvorming praktisch een haalbare kaart is. Een bedrijf heeft nu eenmaal een enorme voorsprong in technische knowhow, hetgeen het voorstellingsvermogen van een individu, maar ook een wetgever, vaak te buiten gaat. Het is dus goed denkbaar dat achter dit gordijn van technische kennis ethische beslissingen worden genomen die men niet als zodanig erkent of herkent. En als een wetgever al iets zou willen reguleren op dit gebied, dan wordt dat vaak belemmerd door de geografische reikwijdte van haar jurisdictie. Plat gezegd: een wetgever kan een landsgrens niet over, een bedrijf wel. En dus 'shoppen' bedrijven tussen landen die de meest gunstige wetgeving op een bepaald gebied hanteren. Dat levert soms spanningen op (denk in een andere context aan de beroemde Panama en Paradise Papers).

Daarnaast is een bedrijf een rechtsvorm met als hoogste doel het maken van winst. Veiligheid of transparantie zijn zeker relevant, maar gaan niet automatisch hand in hand met dat hoogste doel. Het bedrijf heeft er dus niet altijd belang bij de ethische besluitvorming uit te besteden aan hen die het raakt.

En hierin schuilt dus een uitdaging, want het kan misschien wel wenselijk zijn wanneer degene die geraakt wordt door de ethische programmering daar inzicht in heeft en er wellicht invloed op uit kan oefenen.

Ethisch programmeren

Een andere uitdaging zit hem in de ethische opvatting die we kunnen programmeren. In het geval van de zelfrijdende auto willen we die zodanig programmeren dat zo min mogelijk leed wordt berokkend (een soort utilitarisme, gevolgenethiek). Dat zou kunnen betekenen dat de auto ervoor moet kiezen de fietser met helm aan te rijden, omdat zij de grootste kans te overleven heeft. Kiezen we voor een meer principiële benadering (deontologie), dan zouden we kunnen zeggen dat we niet diegene moeten straffen die juist diens verantwoordelijkheid genomen heeft door veilig aan het verkeer deel te nemen met een helm. En wanneer we Kantiaanse principes toepassen op de optie om de inzittenden op te offeren, verliest de auto zijn bestaansrecht – immers: niemand zal de auto dan nog willen aanschaffen. En aangezien we dan helemaal niet aan programmeren toe zullen komen, zou het daarmee dus een zeer onlogische optie zijn.

Deze bovenstaande ethische benaderingen hebben één ding gemeen: ze gaan uit van ofwel gevolgenethiek ofwel beginselenethiek. Dergelijke benaderingen zijn absolutistisch. Dat wil zeggen dat ze leefregels propageren die we allemaal moeten toepassen om een goed leven te kunnen leiden, ongeacht de persoon en ongeacht de situatie. Een soort 'one size fits all'-aanpak. Ideaal om te gebruiken in programmeren dus. Zeker wanneer we technieken als 'deontic logic' gebruiken, waarbij machines geprogrammeerd worden aan de hand van voornamelijk 'verplichte handelingen', 'toegestane handelingen', en 'verboden handelingen' (Von Wright, 1951).

Echter, het wordt een wat complexer verhaal wanneer we ervan uitgaan dat de zelfrijdende auto op dezelfde manier ethiek bedrijft als mensen, dan wel menselijke ethiek imiteert. Hoe populair deze twee stromingen ook zijn in veel literatuur over normatieve ethiek, er zijn best wel wat redenen om aan te nemen dat mensen in het dagelijks leven zelden daadwerkelijk gevolgenethiek of beginselenethiek toepassen. Het is misschien veel realistischer om deze theorieën te zien als hooguit een bron van inspiratie, maar dat ethisch handelen van mensen vooral voortvloeit uit de context waarin mensen zijn opgegroeid, de dingen die men heeft meegemaakt, en de manier waarop men dus een ethisch dilemma herkent en ernaar handelt (Johnson, 1992). Het is goed denkbaar dat mensen patronen van ethisch denken leren wanneer ze opgroeien, en dat dit sterk afhankelijk is van hoe dat er precies uit ziet: dit vormt iemands morele voorstellingsvermogen. Dit morele voorstellingsvermogen is een ervaring, die uniek is, en onmogelijk door een machine kan worden nagemaakt. Ter vergelijking: een machine kun je best programmeren de kleur rood te herkennen en daar een bepaalde handeling aan te koppelen, maar je kunt een machine nooit programmeren hoe het is om de kleur rood te ervaren. Datzelfde geldt voor ethiek: je kunt een machine best programmeren een ethische beslissing te maken op basis van regels, maar niet een ethisch gevoel meegeven op basis van levenservaring.

Conclusie

1. We kunnen machines best inzetten om ethiek te bedrijven, maar nooit om menselijke ethiek te vervangen. Wanneer we doen alsof dat wel kan, ontstaat er een moreel vacuüm.

2. Bedrijven hebben een voorsprong ten aanzien van technologische knowhow, waardoor degene die de gevolgen van ethisch programmeren ondergaat hier niet altijd invloed op kan uitoefenen. Daarnaast is dat ook niet altijd in het belang van het bedrijf. Hier ontstaat ook een moreel vacuüm.

Het is interessant te reflecteren hoe om te gaan met deze morele uitdagingen, waarvan te verwachten valt dat we die we in de toekomst steeds vaker en nadrukkelijker zullen tegenkomen.

Referenties

Johnson, M. (1993). *Moral imagination, Implications of cognitive science for ethics*. Chicago: The University of Chicago Press.

Millar, J. (2017). *Ethics settings for autonomous vehicles*. In: Lin, P., Jenkins, R. & Abney, K. (editors). *Robot Ethics 2.0*. Oxford: Oxford University Press.

Nyholm, S. & Smids, J. (2016). *The Ethics of Accident-Algorithms for Self-Driving Cars: an Applied Trolley Problem?* *Ethical Theory and Moral Practice* (2016) 19:1275–1289.

Wright, G.H. von. (1951). *Deontic logic*. *Mind*, vol. LX, No. 237