

Moreel kompas bij technologische innovaties

Edgar Karssing

Dit verhaal is gebaseerd op de lezing 'Moreel Kompas bij Data-Analytics' die Edgar Karssing heeft gegeven op 10 maart 2020 tijdens het /Podium van DataFryslân.

Inleiding

Data-analytics, kunstmatige intelligentie, het internet of things, robotic process automation. Snelle technologische veranderingen roepen veel ethische vragen op. Wat zijn de voor- en nadelen? Wie is eigenaar van de data en, nog belangrijker, van de algoritmen die voor ons denken? Wie heeft er baat bij, en wie helemaal niet? Wie is waarvoor verantwoordelijk? Hoe komen we tot een moreel kompas? Allemaal vragen waarop nog nauwelijks antwoorden zijn. De discussie staat zeker niet stil. Rapport na rapport verschijnt. Met pleidooien voor het belang van een moreel kompas. Met voorstellen hoe dat kompas eruit zou kunnen zien. Ik zie lijstje na lijstje met belangrijke waarden langskomen, zoals autonomie, menselijke waardigheid, privacy, rechtvaardigheid, solidariteit, veiligheid, duurzaamheid. De uitdaging is de discussies op de juiste plek te voeren: daar waar de technologische innovaties handen en voeten krijgen. Dan moet je de meer filosofische discussies hanteerbaar maken. In deze bijdrage doe ik twee voorstellen waarmee je direct aan de slag kunt. En als je nu denkt: waarom zou ik dat doen? Omdat je geen Victor Frankenstein wilt worden.

Schrikbeeld

De roman *Frankenstein* van Mary Shelley is al meer dan 200 jaar oud. Het boek is geschreven in de begindagen van de industriële revolutie. En is hoogst actueel. De hoofdpersoon, Victor Frankenstein verzucht in het boek: 'Leer van mij, zo niet van mijn aannames dan toch van mijn voorbeeld, hoe gevaarlijk de verwerving van kennis is en hoeveel gelukkiger de man is die meent dat zijn geboortestad de wereld is, dan hij die de ambitie heeft groter te worden dan zijn aard hem toestaat'. Voor alle duidelijkheid, het boek is geen anti-techniekboek. Het biedt een metafoer hoe techniek kan ontsporen als de vereiste zorg niet in acht wordt genomen. Het verhaal is waarschijnlijk bekend, hoewel de meeste verfilmingen geen recht doen aan het boek. Victor creëert met de lichaamsonderdelen van dode mensen een levend wezen. Het wezen gedraagt zich monsterlijk en vermoordt verschillende mensen. Toch heeft dit wezen geen rare wensen, zoals uit een hele mooie dialoog tussen Victor en zijn schepping blijkt: het wil erkenning, liefde en aandacht. Door zijn monsterlijke verschijning is dit het wezen echter niet gegund. Door gebrek aan liefde ontspoot het wezen.

Welke lessen kunnen we uit dit boek halen? Hoezo wil je geen Victor Frankenstein worden? Victor was hoogmoedig – *hubris* – hij zag zichzelf als een god die zijn eigen wezens kan scheppen. Hij had alleen zijn eigen roem en faam voor ogen, niet het verbeteren van welvaart en welzijn van zijn medemensen. Hij nam geen verantwoordelijkheid voor het wezen – het wezen kreeg geen erkenning, liefde en aandacht. Hij nam ook geen verantwoordelijkheid voor de daden van het wezen. Hij wist dat het wezen zijn broertje had vermoord, maar durfde dit niet te zeggen omdat dan bekend zou worden dat hij de schepper was van dit wezen; de au pair van de familie kreeg de schuld – onterecht dus – en werd ter dood veroordeeld. Victor schoot ook tekort doordat hij niet van tevoren nadacht over de toekomst: wat zouden de gevolgen kunnen zijn als zijn 'technologise innovatie' eenmaal in de wereld zou rondlopen? En hij nam al zijn beslissingen in eenzaamheid. Hierdoor ontnam hij zichzelf de gelegenheid tot kritische tegengeluiden. Ook hier geldt: de grootste fout is de fout die je alleen maakt.

Moral Impact Analysis

Kortom, terug naar 2020, wat moet je doen als je geen Frankenstein wilt worden? Neem je verantwoordelijkheid voor je eigen technologische innovaties en zorg dat je met anderen, vooraf, nadenkt over de mogelijke gevolgen van die innovaties. Hoe doe je dat? Door op systematische en gestructureerde wijze alle morele en maatschappelijke zorgen te doordenken en hiermee al in de ontwerpfase rekening mee te houden. Een hulpmiddel is de *Moral Impact Analysis* (MIA, dit idee is ontleend aan Wright 2011). Een MIA bouwt voort op een PIA, een *Privacy Impact Assessment*, zoals bekend geworden door de Algemene Verordening Gegevensbescherming. Om even op te frissen, een PIA 'legt in de eerste plaats de privacyrisico's bloot van nieuwe (projecten en initiatieven) of bestaande verwerkingen van persoonsgegevens en draagt bij aan het vermijden of verminderen van deze privacyrisico's. Op basis van deze PIA wordt op systematische wijze inzichtelijk gemaakt hoe groot de kans is dat de privacy van de betrokken personen van wie gegevens worden verwerkt wordt geschaad, waar deze risico's zich voordoen en welke gevolgen daaraan voor hen verbonden zijn' (NOREA 2015). Je zou kunnen stellen dat een PIA naar een deel kijkt (privacy) en een MIA naar het geheel aan morele en maatschappelijke aspecten van nieuwe technologieën. Dat is zeker verstandig bij het ontwikkelen van nieuwe producten en diensten, tijdens het productontwikkelingsproces. Maar omdat de ontwikkelingen momenteel op de reflectie vooruitlopen, lijkt het me niet onverstandig om ook bestaande producten en diensten aan een MIA te onderwerpen. Een MIA doe je bij voorkeur niet alleen. Het is zelfs een belangrijk doel van een MIA om belanghebbenden hierbij te betrekken om op die manier morele aspecten van nieuwe technologieën te identificeren, te articuleren en te bespreken (blz. 200). Ook kunnen belanghebbenden met oplossingen komen (waarborgen) waar je zelf wellicht nog helemaal niet aan had gedacht. Door stakeholders bij de discussie te betrekken, verklein je ook het risico dat een nieuw product 'ontploft' doordat het onverwacht enorme weerstand oproept. Juist door belanghebbenden te betrekken bouw je ook goodwill bij hen op. Op deze manier is een MIA dus zowel een 'early warning system' als een manier om advies te krijgen (blz. 202).

Het organiseren van dergelijke dialogen is natuurlijk spannend. Je weet op voorhand niet wat er gaat gebeuren, wie wat inbrengt, en of er enige mate van consensus kan worden bereikt. Dat laatste zou ook niet per se het doel moeten zijn. Een MIA is een hulpmiddel om beter inzicht te krijgen in de problematiek, de verantwoordelijkheden die men bereid is te accepteren en een invulling en uitwerking van het eigen morele kompas. Een belangrijk onderdeel van een evenwichtig besluitvormingsproces waarin alle kansen, risico's en zorgen worden meegenomen. Hierbij kan gebruik worden gemaakt van een vragenlijst die houvast biedt bij het inventariseren van morele en maatschappelijke aspecten. Bij bijvoorbeeld de waarde 'rechtvaardigheid' formuleert Wright (2011) de volgende vragen:

- 'Has the project identified all vulnerable groups that may be affected by its undertaking?
- Is the project equitable in its treatment of all groups in society? If not, how could it be made more equitable?
- Does the project confer benefits on some groups but not on others? If so, how is it justified in doing so?
- Do some groups have to pay more than other groups for the same service?
- Is there a fair and just system for addressing project or technology failures with appropriate compensation to affected stakeholders?'

Je begint dus met een aantal waarden en geeft aan welke vragen deze oproepen. Doordat je met waarden begint, heb je meteen een eerste kader voor het beoordelen van antwoorden (draagt dit bij

aan... staat dit op gespannen voet met...). Daarbij wordt niet gestreefd naar een uitputtende opsomming van mogelijke vragen, het zijn startpunten voor het goede gesprek.

Begeleidingsethiek

Als ethicus faciliteer ik geregeld dialogen over handelingsvraagstukken: is het wel of niet moreel verdedigbaar om dit te doen? Mag ik vertrouwelijke informatie gebruiken als ik daarmee een misstand aan de kaak stel? Moet ik deze klant vertellen dat zijn hulpvraag door een concurrent beter kan worden opgepakt? Behoor ik me strikt aan de organisatieregels te houden als dit niet in het belang van mijn medewerker is? Een dergelijke insteek werkt echter niet voor een goede dialoog over technologische innovaties. Het zijn ja/nee-vragen. Wanneer stel je die? Aan het begin, in de ontwerpfase? Maar dan wil je juist nog alle ruimte hebben voor exploreren en is onduidelijk waarop de ja/nee-vraag betrekking heeft. Aan het einde, vlak voordat je met de technologische innovatie de markt op wilt? Dan maak je de discussie wel heel zwaarmoedig: kunnen we door of is alle inspanning voor niets geweest? Een kleine doorbraak in mijn worsteling hiermee kwam door het lezen van het rapport *Aanpak Begeleidingsethiek. Een dialoog over technologie met handelingsperspectief* van ECP/Platform voor InformatieSamenleving (link: <https://ecp.nl/wp-content/uploads/2019/11/060-001-Boek-Aanpak-begeleidingsethiek-240165-binnenwerk-digitaal.pdf>). Zij schrijven bijvoorbeeld:

‘De invulling van techniek-ethiek als het aanvaarden of afwijzen van technologie, plaatst techniek en samenleving tegenover elkaar. Technologie vormt in die benadering een potentiële bedreiging voor de samenleving en het is de verantwoordelijkheid van de ethiek om te bepalen welke techniek toegelaten mag worden en welke niet. Maar dit beeld is onjuist. Techniek en samenleving zijn juist fundamenteel met elkaar verweven. Technologie wordt door mensen ontwikkeld met het oog op een bepaalde rol van die technologie in de samenleving. ... De vraag is in de meeste gevallen niet of een technologie toegelaten moet worden of niet, maar hoe we er op een verantwoorde manier mee om kunnen gaan’ (blz. 10).

Dus geen ja/nee-vragen centraal stellen, maar het begeleiden van technologie in de samenleving. Oftewel, de centrale vraag wordt: hoe? ‘Ze richt zich niet op afwijzen of aanvaarden, maar op het waardenvolle wijze ontwerpen, implementeren en gebruiken van nieuwe technologie’ (blz. 11).

De begeleidingsethiek richt zich niet op algemene vragen, maar altijd op specifieke innovaties. In het rapport worden verschillende cases behandeld. Bijvoorbeeld een eetrobot. Hoe ga je van een algemene ja/nee-vraag naar het bespreken van een specifieke innovatie? ‘In plaats van de vraag “mogen we de zorg voor kwetsbare mensen wel aan machines delegeren?” komt dan de vraag: “is er een manier om de waarden die centraal staan in de zorg als uitgangspunt te nemen bij het ontwikkelen en gebruiken van zorgrobots?”’ (blz. 22). Dit betekent overigens niet dat er nooit nee kan worden gezegd tegen een bepaalde technologische innovatie, daar ligt echter niet op voorhand de focus. ‘Maar uiteindelijk blijft ook “niet” een antwoord op de “hoe”-vraag. Als blijkt dat de technologie niet verenigbaar is met onze waarden, geeft het model als uitkomst dat de technologie beter niet gebruikt kan worden’ (blz. 22).

In het rapport worden drie handelingsopties benoemd om de hoe-vraag te beantwoorden:

- *ethics by design*: in het ontwerp van de technologie rekening houden met morele en maatschappelijke waarden. Zorg bijvoorbeeld dat robotarm afwasbaar is, als blijkt dat mensen deze ‘arm’ ook voor andere doeleinden dan eten gebruiken.

- *ethics in context*: de context waarin de technologie wordt gebruikt zo inrichten dat rekening wordt gehouden met morele en maatschappelijke waarden. Bijvoorbeeld met de introductie van auto's kwamen er stoplichten, zebrapaden en vluchtheuvels.
- *ethics by user*: zorgen voor bewustwording en training benodigde vaardigheden van gebruikers.

Hiermee biedt de begeleidingsethiek niet alleen een duidelijke focus (de hoe-vraag), en een procesaanpak, maar ook handvatten voor concreet handelen.

Tot slot

Frankenstein is een heerlijk boek. Al meer dan twee eeuwen oud, maar nog steeds goed leesbaar. En dus hoogst actueel. De ondertitel van het boek is 'the modern Prometheus'. Prometheus was in de Griekse mythologie de held die het vuur van de goden stal om aan de mens te geven. Het vuur staat hierbij voor de techniek – zo zie je in het logo van de TU Delft nog steeds een vlam. Mary Shelly vergelijkt Victor dus met Prometheus. Maar daarin zit een zekere ironie (vgl. Guston e.a. 2017: xxviii). De naam Prometheus wordt ook wel 'vertaald' als 'de vooruitdenkende'. En dat is precies waar het mis ging met Victor Frankenstein. Hij dacht helemaal niet vooruit! Hij was hoogmoedig, hij nam geen verantwoordelijkheid.

Nieuwe technologische innovaties kunnen veel voor de mensheid gaan betekenen. Qua gezondheid en kwaliteit van leven. We kunnen er wellicht de voedsel- en klimaatproblematiek mee oplossen. Ze maken ons leven leuker en gemakkelijker. Maar technologische innovaties hebben ook schaduwzijden. Die morele en maatschappelijke zorgen moeten we serieus nemen. Om geen Victor Frankenstein te worden. Dat is ook mogelijk. Met behulp van een *Moral Impact Analysis* en door met de begeleidingsethiek te onderzoeken welke handelingsopties er zijn om op een verantwoorde wijze te innoveren.

Edgar Karssing is als universitair hoofddocent beroepsethiek en integriteitsmanagement verbonden aan Nyenrode Business Universiteit. Voor reacties en suggesties: e.karssing@nyenrode.nl

Gebruikte literatuur

Guston, D. e.a. (2017), *Frankenstein. Annotated for scientists, engineers, and creators of all kind*, MIT.

NOREA (2015), *Privacy Impact Assessment (PIA). Introductie, handreiking en vragenlijst*.

Shelly, M. (2015/1818), *Frankenstein*, Veen Klassiek.

Tijink, D. en P. Verbeek (2019), *Aanpak Begeleidingsethiek. Een dialoog over technologie met handelingsperspectief*, ECP/Platform voor InformatieSamenleving.

Wright, D. (2011), 'A framework for the ethical impact assessment of information technology', *Ethics and Information Technology*, Vol. 13 Iss. 3, blz. 199-226.