***Column: Responsible Humanity***

Wordt het tijd voor Responsible Humanity?

Deze vraag stelde ik mijzelf de afgelopen periode vaak rondom de ontwikkelingen van de AI-verordening. In Europa proberen we onder andere met deze wetgeving en verschillende impact assessments handen en voeten te geven aan het begrip “Responsible AI”, of in de minder *sexy* uitdrukking: verantwoordelijke kunstmatige intelligentie. Dit betekent zo ongeveer dat kunstmatige intelligentie technologieën zouden moeten voldoen aan een reeks principes, richtlijnen en praktijken die gericht zijn op onderwerpen als ethiek, veiligheid en betrouwbaarheid, verantwoording, de bescherming van het recht op privacy en andere mensenrechten, eerlijkheid en transparantie. We zijn er maar druk mee om dit begrip verder aan te kleden en te duiden, misschien wel uit een behoefte om grip te willen krijgen op AI, wat in de kern juist ongrijpbaar is.

Maar hoe zit dat dan met de mens? In het 300.000-jarige bestaan van de homo sapiens is er niemand geweest die het begrip “Responsible Humanity” eens onder de loep heeft genomen, terwijl dat in deze tijd waar de wereld zich uitstrekt naar technologische voortuitgang en maatschappelijke verandering, een belangrijk vraagstuk is. Wat mij betreft is het op zijn minst gek dat we geen poging hebben gedaan om de ‘verantwoordelijke mens’ definiëren. Sommige filosofen en ethici geloven dat mensen van nature goed zijn. Nog los van het gegeven dat “goedheid” niet te definiëren is, is “goed zijn” niet hetzelfde als “verantwoordelijk zijn”. ChatGPT 3.5, dat zelf een AI is, antwoordt op de vraag wanneer iemand verantwoordelijk is, het volgende: “Over het algemeen wordt iemand als verantwoordelijk beschouwd wanneer zij controle hebben over hun acties, de gevolgen ervan kunnen overzien en redelijkerwijs kunnen worden aangesproken op hun gedrag binnen de context van ethische, juridisch en sociale normen.” Ingewikkeld hieraan is om de context waarbinnen getoetst moet worden, te definiëren. Dit probleem speelt ook in het geval van Responsible AI, maar toch hebben we dit – in ieder geval gedeeltelijk – in kaders en impact assessments weten te gieten.

Als ik kijk naar het wetgevingsproces rondom AI, denk ik dat een verordening die gaat over verantwoordelijk-mens zijn, nog een brug te ver is. Ik zou daarom dan ook willen pleiten voor de invoering van een

Responsible Humanity Impact Assessment (‘RHIA’) die wordt uitgevoerd door een Responsible AI. Voor deze verantwoordelijke algoritmes zou het niet al te lastig moeten zijn om normen te toetsen aan iets ongrijpbaars als een mens, want we doen het andersom immers ook. Probleem is wel weer dat beslissingen die gemaakt worden door een algoritme over een mens, doorgaans weer door een ander mens moeten worden getoetst om aan de normen van “Responsible AI” te kunnen voldoen. Zo blijft het plan voor een RHIA in de uitvoering steken en kan de mens zich verschuilen achter zijn eigen complexiteit.

Zolang deze vicieuze cirkel niet kan worden doorbroken, kan “Irresponsible Humanity” ondertussen haar gang blijven gaan. Slechte intenties, verknipte morele normen en wankelende juridische kaders blijven aan de orde van de dag. In plaats van je machteloos te voelen over dit toekomstbeeld, is het van belang om je aandacht te vestigen op mogelijke manieren om positieve verandering te bevorderen. ‘Een betere wereld begint bij jezelf’ is niet voor niets een gevleugelde uitspraak.