

Bouwstuk Loge Driehoek 6 maart 2018 van Br. F.H.  
**De relatie tussen Artificiële Intelligentie en Humanisme**

### 1. DE VRAAGSTELLING: EINDE VAN HET HUMANISME?

Rond het jaar 2000 was het bij hele en halve intellectuelen mode om van alles het einde te voorspellen: het einde van de natiestaat en van de ideologieën. Nu, ongeveer twintig jaar later hoef je maar even de krant open te doen, om te ontdekken dat van die voorspellingen weinig is uitgekomen. Nationalisme en fundamentele ideologieën zijn immers springlevend.

De laatste jaren is er een nieuwe golf van 'het einde van' voorspellingen. De Israëliische historicus HARARI voorspelt in zijn laatste boek het einde van het westers humanisme, ten gevolge van de opkomst van Art. Int. De Duitse ICT deskundige Hofstetter schreef onlangs over het einde van de democratie. In mijn Bouwstuk beperk ik me tot het eerste thema omdat het verhaal over het einde van het humanisme de kern van onze beweging raakt. Omdat Harari in zijn wereldwijd vermaarde boeken Homo Sapiens en Homo Deus de enige is die een dergelijk gedurfde voorspelling heeft gedaan, beperk ik me tot deze auteur.

Waarom vindt HARARI dat het Westers Humanisme zijn beste tijd heeft gehad? Daar gaat de vraag aan vooraf wat deze man onder Westers Humanisme verstaat,. Als kernelementen noemt hij: een hoge waardering van de individuele mens, van zijn autonomie, van zijn zelfbewustzijn, zijn geweten en zijn vrije wil en vooral het unieke menselijke vermogen tot ontwikkeling van zijn verstandelijke gaven. Het humanisme heeft verder een lage waardering voor dwingend extern, politiek en religieus gezag. Dwingend extern politiek gezag aanvaardt de humanist alleen conditioneel. Het gezag moet voldoen aan eisen van democratie en rechtstaat. Dwingend extern religieus gezag verwerpt de humanist zonder meer, want het recht van levensbeschouwing is een individueel recht. Ik kan me met deze definitie verenigen.

Harari baseert zijn stelling vooral op het meest recente hersenonderzoek, dat geen enkel empirisch bewijs geeft van het bestaan van genoemde kernelementen als zelfbewustzijn en vrije wil van het fenomeen mens in de humanistische visie. Harari veroordeelt het humanisme niet als normatieve visie, nee de feitelijke grondslagen ervan worden door de moderne wetenschap volgens hem ontkracht. Een tweede basis voor zijn stelling ziet hij in ontwikkelingen van de kunstmatige intelligentie, waardoor de mens het in toenemende mate moet afleggen tegenover kunstmatige vormen van intelligentie buiten hem om. Op termijn kunnen de kunstmatige vormen van intelligentie meer over een mens weten dan hij zelf.

### 2. Eerste bezwaren tegen de stelling van HARARI dat het Westers humanisme zijn beste tijd heeft gehad.

De stelling van Harari roept bij mij vooral ongeloof en scepsis op. Zeker als Vrijmetselaar, als lid van een van de oudste humanistische verenigingen ter wereld, kan ik zijn voorspelling niet onmiddellijk omarmen. De stelling roept een aantal vragen op, waar HARARI volgens mij ten onrechte aan voorbij gaat. Bovendien bevat zijn redenering volgens mij een aantal denkfouten. Hierover gaat deze tweede paragraaf van mijn bouwstuk. Om een nog scherper beeld te kunnen krijgen ga ik daarna in de derde paragraaf na welk problemen de mensen die nu daadwerkelijk werken aan de ontwikkeling van robots en van systemen van

kunstmatige intelligentie (ART.Int.) in hun werk ervaren. Wat vinden de mensen, die met hun poten in de modder staan, zelf van de problematiek, die HARARI aansnijdt? Daarna wordt in de vierde en laatste paragraaf definitief geconcludeerd.

Mijn principiële bezwaren tegen het idee van het einde van het Westers humanisme kunnen als volgt worden samengevat SI NON E VERO EBENE TROVATO.

- Allereerst wie, zoals Harari stevige stellingen poneert, heeft veel uit te leggen. Harari doet dat niet. Het westers humanisme ligt niet alleen ten grondslag aan onze Westerse Beschaving, maar in belangrijke mate ook aan onze Wereldbeschaving. Denk in dit laatste verband aan de V.N. verdragen betreffende de individuele mensenrechten. Ook veel maatschappelijke deelsystemen hebben het Westers humanisme als één fundamentele grondslag. Denk aan de politieke democratie op staatkundig terrein en de consumentensoevereiniteit in de economie. De individuele kiezer en de individuele klant beslissen, dat is onze algemeen uitgangspunt op resp. het terrein van de politiek en de economie. Bij ons onderwijs gaat het toch vooral om het primaat van de individuele ontplooiing en om de ontwikkeling van het vermogen om zelf te denken en om een individueel geweten, een moreel kompas te vormen. In het recht worden mensen die niet lijden aan een evidente ziekelijke stoornis van hun geestesvermogen persoonlijk verantwoordelijk gehouden voor de gevolgen van hun gedrag. Als het humanisme een dergelijk kernelement is in onze visie op mens en maatschappij, dan mag volgens mij van een criticus als HARARI op zijn minst verwacht worden dat hij aangeeft, welke gelijkwaardige alternatieven er zijn, wanneer het humanisme volgens hem zal verdwijnen. Dat laat HARARI na. Ook hier geldt het oude adagium: gooi nooit oude schoenen weg, voordat je nieuwe hebt, zelfs al zijn de schoenen, die je aan hebt, wat sleets.
- Een tweede bezwaar betreft de mensvisie. Al die mensen, zoals HARARI, die spreken over de komende superioriteit van de Art.Int. boven de mens, baseren die stelling voornamelijk op de recente overwinningen van met Art. Int. toegeruste Robots bij spelen als schaak en het Chinese Go. Die mensen, die zo denken, hebben volgens mij niet in de gaten dat de mens meer is dan berekenende intelligentie. Harari c.s. gaan er m.i. geheel ten onrechte vanuit dat de meerwaarde van de mens alleen in zijn berekenende intelligentie ligt. Het feit dat de mens het op dat punt steeds meer van vormen van Art Int verliest, zou zijn komende inferioriteit aankondigen. De Duits-Amerikaanse filosoof en ICT-deskundige Weizenbaum, een van de grondleggers van Art.int. is een van de weinigen in de wereld van ICT-deskundigen die zich tegen deze visie verzet. In zijn boek Computer Power and Human Reason wijst hij op de unieke onvervangbare geestelijke capaciteiten van de mens. Hij noemt de wijsheid van de mens, zijn beoordelings-vermogen, zijn vermogen om rekening te houden met kwalitatieve factoren als emoties, deugden en compassie. Computers kunnen volgens Weizenbaum mensen niet vervangen, maar zijn complementair. De mens is ook een sociaal en ethisch wezen en dat leidt tot een voor de mens specifieke intelligentie, die superieur is aan de kunstmatige intelligentie, aldus Weizenbaum.
- HARIRI heeft ten derde onvoldoend oog voor het feit dat alle nieuwe vormen van Art. Int. geheel eigen ethische, normatieve vragen oproepen, waarop de technologie zelf en het

denken in termen van doelmatigheid geen antwoord hebben. Een goed voorbeeld vormen de zelfrijdende auto's. Technisch zijn ze af, maar de auto-industrie zit in haar maag met enkele moeilijke vragen van ethische, normatieve aard. Hoe moet de software van een dergelijk auto geprogrammeerd worden? Mag in de software van de auto bij botsing worden vastgelegd dat het belang van de inzittende altijd voorgaat, ook als dit de dood van andere weggebruikers betekent? Hoe relevant zijn hier de schuldvraag of de leeftijd van de bedreigden? In deze situatie waarin de nieuwe technologie veel normatieve, ethische vragen oproept, is het humanisme als set van voorwaarden volgens mij nog steeds bruikbaar om bouwstenen voor een normatief beoordelingskader van de toekomstige technologie te leveren. De wereldwijde, Europese aanvaarding van het humanisme is hierbij het sterke punt, om te komen tot wereldwijd, subsidiair op Europees niveau geldende normen.

#### 4. De visie van de mensen uit de praktijk

Laten we vervolgens eens kijken wat mensen vinden die als onderzoeker zelf werken aan de ontwikkeling van systemen van Art. Int., mensen, die dus zelf met de poten in de modder staan. Wat zeggen zij over de gewenste verdere ontwikkeling van Art.Int. ?

De TNO onderzoeker Bronkhorst (V.16 nov.) stelt dat: de beide uitersten in onze kijk op Art.Int. - als brenger van de heilsboodschap of als dreigend, af te weren doembeeld, niet aan de orde zijn. In de fase van de huidige ontwikkeling gaat het er volgens hem allereerst om dat we de zwakke punten van de huidige vormen van Art. Int. wegwerken, namelijk hun onvermogen om met de mens samen te werken en hun ongevoeligheid voor de menselijke doelen, behoeften en intenties. In de huidige tijd moeten wij ons vooral inzetten om vormen van Art Int., zoals robots, **sociale** intelligentie bij te brengen, zodat robots zich gedragen als vriend, vraagbaak en ondersteuner van de mens. Een ander zwak punt van de huidige robots en daarmee verbonden vormen van Art Int. is volgens Bronkhorst dat zij alleen succesvol zijn in strak gestructureerde en begrensde omgevingen, zoals een schaakbord. Dit betekent dat intelligente robots braaf conform hun opdracht adequaat kunnen reageren op voorgeprogrammeerde situaties. Deze robots zijn echter in hun reactie onvoorspelbaar en gevaarlijk in situaties, die de programmeur niet had voorzien. Ook dat zwakke punt moet eerst worden weg gewerkt.

Het meest essentiële gebrek betreft volgens vele mensen, die in die sector werken, echter de verwaarlozing van de ethische component in de huidige ontwikkelingsfase. De Delftse hoogleraar Jonker, specialist in het leren van robots, wijst erop dat robots van zich zelf ethisch neutraal zijn. Alles hangt af van de ethische intenties van de mens die de robot leert. Ook Jonker vindt dat voorrang moet worden gegeven aan het meer dienstvaardiger, ethischer maken van robots. Het heeft geen zin de robot nog slimmer te maken in situaties waarin hij al goed is, zoals bordspelen. Ten slotte, een opmerking over een van de meest gevaarlijke vormen van kunstmatige intelligentie, de zgn. killing robots of drones. De Ned. deskundige op het terrein van killer robots mw. Verdriesen zegt dat niemand weet of en welke ethische normen zijn verwerkt in de software en algoritmes van de huidige, al operationele killerrobots. Zij erkent dat de aandacht voor dit ethisch aspect zelfs op dit terrein van dood en leven de afgelopen jaren flink verwaarloosd is en pas sinds kort de aandacht heeft van overheden, wetenschappers en bedrijfsleven. Googles dochter DEEP MIND heeft pas sinds enkele jaren een ethische afdeling. Wereldwijd wordt gepleit voor een absoluut verbod van de killerrobots.

De Amsterdamse hoogleraar WELLING (NRC 25. NOV.) wijst vooral op de grote maatschappelijke problemen die Art. Int. oproept. Is bij voorbeeld de huidige solidariteit in de zorg nog wel mogelijk als voor ieder voorspelbaar is welke vreselijk ziektes hij zal oplopen. Hoe is het verzamelen van miljoenen data van mensen te verenigen met steeds scherpere privacyregels van de overheid? Wat is de rol van Europa, nu Chinese en Amerikaanse tech-reuzen in deze sector zo domineren? WELLING suggereert dat Europa zich maar het best kan richten op een, ook volgens hem verwaarloosd aspect, te weten algemene vragen van ethiek. Het Europees beleid gaat overigens al in die richting. De E.U. heeft onlangs een advies van de Europese SER overgenomen, waarin het toezicht op vooral de ethische en juridische aspecten van kunstmatige intelligentie centraal staat. Twee middelen hiertoe worden voorgesteld. Er moet een stelsel van normen en certificaten komen dat garandeert dat alle apparaten in deze sector voldoen aan eisen van veiligheid, transparantie, begrijpelijkheid, verklaarbaarheid en ethische en juridische normen. Ook moet in ieder apparaat een kill switch worden ingebouwd. Een kill switch is een druk op de knop die in noodgevallen alle functies van het apparaat lamlegt, als het ongehoorzaam is, zijn eigen weg gaat en schade aanricht.

## 5. Conclusies

Wat leert dit kort overzicht van het denken in de praktijk over robots en art. I? Waar kamergeleerden als HARARI zich druk maken over ICT en het einde van de mens en van het humanisme voorspellen, blijken in de huidige reële ontwikkelingsfase van Art Int. heel andere thema's urgent. Die liggen vooral in de zwakke maatschappelijke en ethische conditionering en regulering van Art. Int. Zolang die zwakke kanten niet zijn verdwenen, kun je m.i. in de huidige ontwikkelingsfase van Art Int. in het geheel niet spreken van een volwassen technologie die voldoet aan eisen van bruikbaarheid, mensvriendelijkheid en betrouwbaarheid. Juist de zorg voor een goede ethische component van de Art Int. vraagt gebruik van belangrijke ethische bronnen, die door de eeuwen heen en wereldwijd hun waarde hebben bewezen, zoals het humanisme. Kortom er is geen sprake van het einde van het humanisme. Integendeel humanistische ethiek hebben we in de toekomst juist heel hard nodig. (HAAS / SCHILDPAD ICARUS)

## 6. Tenslotte

Het bovenstaande betreft alleen maar deelwaarheden en ontzenuwde onwaarheden op onderdelen. Bestaat er ook een bevredigende *alomvattende visie* op deze complexe thematiek? Aanzetten daartoe vond ik vooral in de jonge tak van wetenschap, van de filosofie en de ethiek van de techniek.

Kernpunt hierin is dat we heel anders moeten gaan denken over de relatie tussen mens en techniek. Rode draad in het traditionele westerse denken is een scherpe scheiding tussen de mens en de techniek. Die scheiding heeft oude, vooral religieuze wortels. Het christendom zet de mens immers op een voetstuk. Alleen de mens heeft een ziel, vrije wil, zelfbewustheid enz. Bekend verhaal. Die traditionele scheiding moet volgens de huidige technologische ethici plaats maken voor een visie, **waarin de onderlinge verwevenheid tussen mens en techniek voorop** staat. De mens is immers het enige wezen dat gebruik kan maken van technische instrumenten zoals kooktechnieken, werktuigen enz. Daarnaast is veel ook

bestaande techniek zelf impliciet gebaseerd op ethische keuzes. Techniek moet immers veilig zijn omdat wij ons leven en dat van anderen willen sparen. Andersom beïnvloeden nieuwe technieken het menselijk gedrag en moraal. Voorbeeld. De pil voor geboortebepanking verbrak de vroegere strenge band tussen seks en voortplanting.

Daardoor krijgt de huidige technologische ethicus ook een andere plaats. Waar vroeger de ethicus alleen of met collega's op zijn kamertje zat te filosoferen, is de moderne technische ethicus vanaf het begin actief lid van het team dat nieuwe technologische constructies en producten ontwikkelt.

0-0-0-0-0