

Nexus Mensen, Informatie, AI



Petra Bolhuis
Maieutiek.nl

1

Wat gaan we tegemoet?



‘Beschavingen worden geboren uit huwelijken tussen bureaucratie en mythologie’ (zie college 4). ‘Het computernetwerk is een nieuw soort bureaucratie die veel krachtiger en genadelozer is dan alle menselijke bureaucratieën die we ooit hebben gezien. Dit netwerk zal vermoedelijk ook intercomputermythologieën creëren die veel complexer en vreemdsoortiger zullen zijn dan alle geoden die mensen ooit verzonnen hebben.’ Dit zegt Harari over hoe hij de toekomst ziet (blz 349).

Zijn bezorgdheid is niet dat we dit niet overleven als mensheid, ook al is dat ook een optie. Zijn bezorgdheid is dat we veel ellende zullen moeten doorstaan, omdat we niet op tijd begrijpen waar en hoe we de menselijke waarden moeten verdedigen tegen de nieuwe techniek die a-moreel handelt en het menselijke van de mens er niet of nauwelijks in geprogrammeerd krijgt.

2

‘Technologie leidt vaak tot historische rampen’ (350)



- Imperialisme, industrialisatie en andere grote sprongen in de ontwikkeling van de mensheid hebben een spoor van ellende getrokken en de mensheid is nog steeds bezig de onbedoelde en onvoorziene gevolgen daarvan te corrigeren.
- Harari benadrukt de ellende van de nieuwe techniek, maar wat hij ellende noemt is eeuwenlang vooral beschreven als vooruitgang.
- Er is sprake van een paradox. Dat is goed te zien aan het concept slavernij. Is dat, zoals Harari schrijft, het gevolg van technische vooruitgang, of hebben we slavernij juist afgeschaft als gevolg van de technische vooruitgang, zoals anderen claimen? (Robin Blackburn en Eric Williams)
- De combinatie techniek en moraal staat steeds op gespannen voet met elkaar.

Wat zijn reële en actuele problemen?



De vraag is of AI inter-computer mythologieën gaat bedenken en als dat gebeurt of daar ons grootste probleem ligt.

AI is een informatienetwerk dat informatie verdeelt en verbindt. Mensen die baat hebben bij bepaalde verhalen kunnen deze sturen naar veel andere mensen. Sommigen zullen een monopolie verkrijgen op informatie en verspreiding. Dat is een zeker en actueel gevaar.

Daarnaast is er het probleem van de toegang tot energiebronnen, water en grondstoffen die nodig zijn voor alle AI. Ook daar is een actueel bestaand gevaar.

Kijk hoe Trump mensen toegang tot hun onderzoeksdata afneemt. Je kunt geen dossiers meer mee naar buiten smokkelen in dit computertijdperk. Je AI- gegenereerde toegang is per direct opgeheven.

De lus tussen moraal en techniek



- Achterhuis, Verbeek en Jos de Mul zijn een paar van de hedendaagse filosofen die veel hebben nagedacht over moraal en techniek. Het grootste probleem is, en dat lijkt ook bij Harari een rol te spelen, dat pas wanneer de techniek al werkzaam is grenzen aan de techniek worden opgelegd door morele overwegingen. Die ethische discussie wordt echter zelf gekleurd door de nieuwe technieken.
- Wanneer machines de arbeid van mensen over kan nemen, pas dan ontstaat de discussie over wat er van arbeiders gevraagd mag worden. Sterke morele overtuigingen gedragen door een groep en voorzien van een krachtige metafysica veranderen het laatste. Maar ook de Amis hebben een noodtelefoon en ook de SGP en de Antroposofie gebruiken anno 2025 computers.
- Wij mensen veranderen dus in onze opvattingen door de technische mogelijkheden. Maar daar schuilt een probleem: er is dus geen werkelijk buitenperspectief. De techniek van computers en AI maakt, beheert en distribueert informatiestromen, waardoor waarden kunnen verschuiven. In tegenstelling tot de wereld vóór AI, is de ruimte voor kritiek en dialoog nu zelf in handen van de technische verandering waarover wij mensen ons een oordeel moeten vormen.

Een paar van de problemen



1. **Bias en discriminatie:** AI kan bestaande ongelijkheden en vooroordelen versterken.
2. **Transparantie en uitlegbaarheid:** veel modellen zijn "black boxes", waardoor beslissingen moeilijk te controleren zijn. De informatie is domweg teveel om allemaal te controleren. Harari waarschuwt dat we AI meer waarheid toekennen dan veilig is. Niet nieuw want dit deden we ooit ook met leiders en teksten of wetenschappers. (bijv. Lombroso)
3. **Kwetsbaarheid:** AI-systemen kunnen gemanipuleerd of misleid worden. (adversarial attacks). Opzettelijke misleiding van de verkeersbordherkenning in een auto. Ouderwets jammen van informatie waarop piloten zich moeten baseren is ook heel disruptief. De vraag is of het de techniek is, of de moraal, die hier het probleem is.



4. **Privacy en surveillance:** Massale data-analyse leidt tot zorgen over verlies van privacy. Data aanlevering is overal en vaak onvermijdelijk. Gezichtsherkenning en trackingtechnologie kunnen leiden tot ongewenste surveillance. Wie krijgt straks toegang tot deze techniek en de bijbehorende data?
5. **Desinformatie en manipulatie** AI-gegenereerde content (deepfakes, synthetische teksten/beelden) maakt het moeilijk feit en fictie te onderscheiden. Er is (nog) geen wet om daartegen actie te ondernemen, ook al vindt iedereen het onwenselijk. Het kan worden ingezet voor politieke beïnvloeding, propaganda of fraude.



6. **Veiligheids- en militaire risico's:** Autonome wapens: AI in drones en wapensystemen roept ethische en veiligheidsvragen op. Cybersecurity: AI kan aanvallen automatiseren en krachtiger maken.
7. **Lange-termijn en existentiële risico's:** Controleprobleem: de mogelijkheid dat zeer geavanceerde AI-systemen doelen nastreven die niet in lijn zijn met menselijke waarden. Existentiële bedreiging: de zorg dat een "superintelligente" AI (AGI) de mensheid kan overheersen of vernietigen. Dit wordt serieuzer genomen in academische literatuur sinds o.a. Nick Bostrom en rapporten van OpenAI/DeepMind.



8. **Economische en maatschappelijke gevolgen:** Bij voorbeeld voor de arbeidsmarkt: vervanging van banen, nu vooral routinematig- of administratief werk, maar de verwachting is dat het daar niet stopt.
9. **Concentratie van macht:** Grote Tech bedrijven hebben disproportioneel veel controle en data en de democratische politieke invloed is klein.
10. **Afhankelijkheid van technologie:** risico van verlies van menselijke vaardigheden en autonomie.
 - Denk aan Scherders 'Afnemende Intelligentie' en de term 'Breinrot'. De term 'brain rot' is van Henry David Thoreau in zijn boek *Walden* uit 1854!
 - Maar ook kunnen mensen alleen via de computer bij meer en meer data. Bijv. onderzoeksresultaten, identiteitscontrole en financiën bestaan alleen nog op een computer/ in the cloud.

Einde van de menselijke geschiedenis?



- Harari zegt op blz 252 al: “we hebben het hier over het mogelijke einde van de menselijke geschiedenis.” Onder deze bewering liggen een aantal aannames die Harari vervolgens uitlegt:
- “We leven in een cocon en ervaren de realiteit door een cultureel prisma.”
- “Mensen hebben millennia lang in de dromen van andere mensen geleefd.”
- In tegenstelling tot de film *The Matrix* hoeven mensen niet fysiek geïsoleerd en met een computer verbonden te zijn om onder invloed van dit AI verhaal te raken.
- AI maakt Plato's grot en Descartes demon (die je i.t.t. de Goede God misleidt) mogelijk.

De voorbeelden van eerdere grote systemen die zich beschermden



- **Wetenschap** is in zich zelf een manier om de toenemende kennis te organiseren, maar dit systeem staat nu zelf onder druk omdat het werkt binnen een kleine groep mensen die onder gecontroleerde omstandigheden kennis verzamelt, verbindt en distribueert.
- Dennett (389) zoekt het in de **financiële markt** die wetten tegen valsmunterij invoerde waardoor het systeem stabiel bleef. Daarbij is er steeds een financiële reserve om het onvoorziene het hoofd te bieden. Maar ook dit staat onder druk door het gebrek aan bodem in de realiteit De goud- of andere reserves worden losgelaten en er is een de toenemend aandeel virtueel geld (bitcoins).

Nexus (knooppunt in een netwerk)



- Harari legt uit dat hij het heeft over het AI netwerk omdat hij het tegenover het menselijke netwerk zet.

De vraag is wie er in het centrale punt (nexus) zit. Er zijn twee gevaren:

1. Er zit **niemand** en dus gaat het netwerk een eigen en ongecontroleerd bestaan leiden. Waarbij het niet per se gericht is op menselijke welbevinden of overleven.
2. Er zijn **mensen met macht** en die beheersen het netwerk beter dan de andere deelnemers en de controleurs. Die weten dat veelal niet en als ze het wel weten hebben ze geen macht om het te veranderen. Informatieasymmetrie noemt Audrey Tang dat (266-267).

Conclusies van Harari (blz 450):



- Menselijke informatienetwerken stellen vaak orde boven waarheid dat leverde veel macht op, maar weinig waarheid.
- Naar mate informatiesystemen krachtig worden kunnen ingebeelde angsten zelf meer rampen veroorzaken dan natuurrampen.
- Zelfcorrectie wordt steeds belangrijker naarmate de techniek meer kan.
- De wet van de jungle is net zo min waar als een grote vooruitgang richting eeuwige vrede.
- Homo sapiens is niet voorwaardelijk voor de natuur. Mensen zijn vermoedelijk het enige stukje van de kosmos dat zelfbewustzijn heeft daarmee hebben wij de verantwoordelijkheid om dat te laten voortbestaan.
- Elke moreel-ethische keuze is een menselijke keuze, dus ook wat AI uiteindelijk gaat doen. Onderdeel daarvan is de ruimte voor zelfcorrectie die mensen nodig hebben ook te implementeren in AI systemen.