Mühlhoff, Rainer, Künstliche Intelligenz und der neue Faschismus, Stuttgart 2026

4. De ideologieën achter de populaire AI-hype

Populaire AI-verhalen zijn gebaseerd op drie fundamentele elementen: de antropomorfisering van AI-systemen, emotioneel overdreven verwachtingen van de toekomst (zowel utopisch als apocalyptisch) en technologische beloftes op korte termijn van oplossingen voor maatschappelijke problemen (solutionisme). Deze ideeën wijken vaak aanzienlijk af van de werkelijke stand van de AI-technologie. Maar wat schuilt er achter deze hype die het potentieel van AI-technologie zo drastisch overdrijft?
Trappen de politiek en het publiek simpelweg in de val van de economische belangen van de AI-industrie?

Hieronder zal duidelijk worden dat het niet zo eenvoudig is. Hoewel de publieke AI-hype inderdaad grotendeels draait om economische belangen, wordt de vermeende overschatting van AI gevoed door langdurige intellectuele stromingen die diep geworteld zijn in wetenschappelijke en industriële kringen, en die samengevat kunnen worden onder de overkoepelende term '*technologische ideologieën*'. Deze stromingen omvatten technologisch determinisme in zijn verschillende vormen, transhumanisme en bepaalde futuristische interpretaties van utilitaristische ethiek – met name longtermisme. Deze stromingen bieden niet alleen argumentatieve rechtvaardigingen, maar beïnvloeden ook de perceptie- en evaluatiepatronen waardoor AI als een historisch unieke, op zichzelf staande macht verschijnt.

Door deze verschillende stromingen gezamenlijk te beschrijven en te analyseren als technologische ideologieën, willen we aantonen hoe de wederzijdse versterking van industriële belangen en de publieke perceptie van AI de machtsbelangen die met AI verbonden zijn, verhult achter quasi-natuurlijke waarheden. Aan de hand van het uienschilmodel dat aan het begin van het vorige hoofdstuk werd beschreven, bewegen we ons nu van het oppervlak van het publieke debat naar de eerste diepere lagen van de complexe en heterogene ideologische constructie die we tegenkomen in de wisselwerking tussen AI-technologie en fascistische politiek. Zoals eerder aangekondigd, worden de meeste ideologische verhalen over AI in het publieke debat ondersteund door verbindingen met diepere ideologische lagen – en het zijn juist deze verbindingen die nu onderzocht moeten worden.

85-86

Technologisch determinisme en techno-optimisme

Alle hieronder beschreven technologische ideologieën zijn in eerste instantie verbonden met de agressieve innovatiecultuur die met name in Silicon Valley te vinden is.[1] Deze cultuur is nauw verweven met een ondernemend durfkapitalisme dat door de private sector gedreven technologische 'disruptie' ziet als de drijvende kracht achter maatschappelijke vooruitgang. Tegelijkertijd wordt ze gekenmerkt door een *libertaire ideologie* die overheidsregulering als een obstakel voor innovatie beschouwt en bijna onbeperkte controle toekent aan de economische – en tevens de sociale – selectiefunctie van de markt. Bovendien zorgt de nauwe samenhang tussen academisch onderzoek en technologische ontwikkeling in de private sector ervoor dat economische belangen en ondernemend denken het wetenschappelijke discours sterk beïnvloeden. In deze omgeving worden technofiele maatschappelijke utopieën in het bijzonder gepromoot en verspreid.

De opvatting dat technologische ontwikkeling een autonome kracht is die de maatschappelijke structuren, de vooruitgang en de loop van de menselijke geschiedenis significant en onvermijdelijk bepaalt, wordt vaak aangeduid als *technologisch determinisme*. Een prominent voorstander van deze visie is Sam Altman, CEO van OpenAI, die herhaaldelijk verklaart dat de AI-revolutie “niet te stoppen" is en onvermijdelijk "fenomenale rijkdom" zal genereren.[2] Deze veel geciteerde uitspraak uit 2015 is ook van hem afkomstig: "Ik denk dat AI waarschijnlijk, hoogstwaarschijnlijk, zal leiden tot het einde van de wereld. Maar ondertussen zullen er geweldige bedrijven zijn."[3] Hij laat zien hoe technologisch determinisme gemakkelijk te combineren is met de perspectieven van economisch en sociaal libertarisme. Dit creëert een visie waarin startups verschijnen als de wereldse actoren en manifestaties van de natuurlijke kracht van technologische ontwikkeling. De puur descriptieve theorie van technologisch determinisme kan zo gemakkelijk worden gecombineerd met een normatieve legitimatie van onbeperkte ondernemersvrijheid.

Een andere profeet van een dergelijk normatief en visionair georiënteerd technologisch determinisme is de Silicon Valley-investeringsbankier Marc Andreessen met zijn invloedrijke *Techno-Optimistisch Manifest (The Techno-Optimistic Manifesto)* uit 2023. Daarin staat:

Onze beschaving is gebouwd op technologie. Technologie is de glorie van menselijke ambitie en prestatie, de speerpunt van vooruitgang en de verwezenlijking van ons potentieel.
Honderden jaren lang hebben we dit terecht verheerlijkt – tot voor kort. Ik ben hier om het goede nieuws te brengen. We kunnen een veel betere manier van leven en zijn bereiken.[4]

De techno-optimisten in Silicon Valley verkondigen met klem de overtuiging dat technologische vooruitgang niet alleen onvermijdelijk, maar ook inherent wenselijk en moreel noodzakelijk is – maar dat deze momenteel wordt belemmerd door liberaal en op duurzaamheid gericht beleid. Ze verwerpen regelgevende interventies als vijandig tegenover innovatie en promoten een wereldbeeld waarin technologische 'ontwrichting' de bron is van welvaart, maatschappelijke vooruitgang en een paradijselijke toekomst.

Een notoire omissie in deze denkwijzen is het onderzoek naar vragen over de maatschappelijke verdeling van macht en middelen, of de reeds zichtbare maatschappelijke en milieuschade veroorzaakt door AI-technologie. Het feit dat de AI-industrie zelf de voornaamste begunstigde is van deze techno-deterministische en techno-optimistische verhalen wordt eveneens genegeerd. Tegelijkertijd zijn deze verhalen zo opgebouwd dat de protagonisten de vraag naar een eerlijke verdeling van de rijkdom die de samenleving als geheel heeft vergaard, kunnen ontwijken. Een centraal aspect van het ideologische karakter van deze wereldvisie is dat deze punten systematisch worden verhuld.[5]

86-89

Transhumanisme en zijn onderstromen

In een invloedrijk kritisch artikel uit 2024 introduceren computerwetenschapper, K1-ethicus en voormalig Google-medewerker Timnit Gebru en de Amerikaanse filosoof Émile Torres het acroniem TESCRAL om een ‘bundel’ van met elkaar verweven en overlappende ideologieën en doctrines te beschrijven die wijdverspreid zijn in bepaalde technologische en wetenschappelijke kringen, vooral in Silicon Valley.
De afkorting TESCRAL staat voor ‘Transhumanisme, Extropianisme, Singularitarianisme, Kosmisme, Rationalisme, Effectief Altruïsme en Langetermijnisme’. (Transhumanism, Extropianism, Singularitarianism, Cosmism, Rationalism, Effective Altruism, and Longtermism). Hoewel het geen gesloten ideologie is, maar eerder een losse combinatie van ideeën die deels tientallen jaren oud en deels heel nieuw zijn, stellen Gebru en Torres dat deze ideologische lessen de verhalen van grote technologiebedrijven, invloedrijke investeerders en opinieleiders in AI-onderzoek aanzienlijk bepalen. Als ideologieën brengen ze een ontologische en ethische oriëntatie over die toekomstige technologieën zoals AI, nanotechnologie en genetische manipulatie niet alleen als onvermijdelijk positioneert, maar ook als moreel wenselijk en noodzakelijk voor het voortbestaan van de mensheid op de lange termijn. Ze dienen ook als ideologische rechtvaardiging voor de technologische en geopolitieke strategieën van talloze bedrijven en actoren op dit gebied. Het blijkt ook dat sommige van deze wereldbeelden een

directe voortzetting zijn van de eugenetica en rassentheorie van de 20e eeuw.
Transhumanisme vormt een van de centrale ideologische wortels van de TESCREAL-ideologieën. De oorsprong ervan gaat terug tot het midden van de 20e eeuw. Het is gebaseerd op de overtuiging dat mensen door het gerichte gebruik van technologie hun fysieke, biologische en intellectuele grenzen kunnen overwinnen, kunnen opgaan in een hybride biologisch-technologische bestaansvorm en daardoor een verbeterd bestaan kunnen bereiken dat vooral ziekte en sterfte zal overwinnen.
De term transhumanisme gaat terug op een essay uit 1957 van de Britse eugeneticus en gedragsbioloog Julian Huxley (1887-1975), die tevens de eerste directeur van UNESCO was (1946-48) en voorzitter van de British Eugenics Society (1937-44, 1959-62). Daar beschrijft hij het idee van transhumanisme als een proces waardoor ‘de menselijke soort als geheel, als mensheid’ zichzelf overwint.[8]
Enerzijds wordt deze groei buiten onszelf beschrijvend beschreven als een evolutionaire ontwikkeling die niet te stoppen is in de zin van technologisch determinisme. Technologische ontwikkeling wordt gezien als een natuurlijke kracht die de menselijke evolutie bevordert. Aan de andere kant kent het transhumanisme in zijn verschillende onderstromen ook een actieve en co-creërende rol in dit proces toe aan mensen als ontwikkelaars van technologie in verschillende betekenissen: in de jaren negentig leidde de Britse futurist Max More bijvoorbeeld de *verplichting* van de mensheid om vooruitgang te boeken af als de morele imperatief van het transhumanisme.[9] Deze zinsnede, die al lang daarvoor vorm had gegeven aan het basisidee van de eugenetica en aan de ‘rassenhygiëne’ van de nationaal-socialisten, toont een kenmerkend keerpunt aan van een normatief en politiek beleid. programma. Dit keerpunt is ook te zien in het Techno-Optimistisch Manifest. Aan de ene kant zijn er sinds het midden van de 20e eeuw, als onderstromen van het transhumanisme, direct bruikbare projecten ontstaan, zoals de ontwikkeling van prothesen om de functie van het menselijk lichaam uit te breiden of te behouden, maar ook nanotechnologie- en bio-engineeringprojecten, bijvoorbeeld in de regeneratieve geneeskunde. Aan de andere kant propageert het transhumanisme duidelijk bizarre tendensen die bijvoorbeeld gericht zijn op de ontwikkeling van brein-computerinterfaces waarmee de menselijke geest of de inhoud van zijn bewustzijn digitaal kan worden ‘gelezen’ en naar een cloud kan worden geüpload (Elon Musks bedrijf Neuralink heeft zich bijvoorbeeld aan dit doel gecommitteerd).
Een centrale actor in de hedendaagse transhumanistische beweging is de Zweedse filosoof Nick Bostrom, wiens werk op het gebied van kunstmatige superintelligentie, ‘existentiële risico’s’ en ‘menselijke verbeteringsethiek’ tot ver buiten academische kringen wordt erkend. Bostrom richtte in 1998 samen met David Pearce de transhumanistische denktank World Transhumanist Association (nu Humanity+) op en leidde tot de ontbinding in april 2024 het Future of Humanity Institute aan de Universiteit van Oxford,

dat werd ondersteund door financiële steun van Silicon Valley en technologiemiljardairs als Elon Musk en Peter Thiel. In zijn boek *Superintelligence: Paths, Dangers, Strategies,* 2014 betoogt Bostrom dat de ontwikkeling van kunstmatige superintelligentie ofwel de ultieme vooruitgang van de mensheid naar een nieuw stadium zou betekenen, ofwel tot haar volledige vernietiging zou leiden. Deze utopisch-apocalyptische overdrijving is hierboven al besproken; ze is typerend voor het hedendaagse transhumanistische denken en is van daaruit overgenomen in het populaire AI-discours.

89-92

Extropianisme en singularitarisme

De eerdergenoemde Max More is sinds de jaren 80 een voorstander van een transhumanistische substroming, bekend als "Extropianisme" (wat ook de oorsprong is van de tweede letter van het acroniem TESCREAL van Gebrus en Torres). De naam is afgeleid van het neologisme "extropie", een antoniem van de fysische grootheid entropie. Extropie duidt volgens More op "de mate van intelligentie, functionele orde, vitaliteit en capaciteit en drang tot verbetering van een levend of organisatorisch systeem."[10] Extropianisme ziet zichzelf als een waardesysteem dat zich toelegt op het maximaliseren van deze extropie en streeft er daarom naar om *proactief in te grijpen in de menselijke evolutie,* bijvoorbeeld door middel van bio-engineering of de samensmelting van mensen met AI-technologie.

Met de term "extropie" neemt het extropianisme een bijna uitdagende houding aan ten opzichte van de tweede wet van de thermodynamica in de natuurkunde. Die stelt dat de entropie (de mate van wanorde) van een gesloten systeem nooit afneemt in de loop der tijd en hoogstens toeneemt – alles evolueert van orde naar chaos. Als echter de entropie van een subsysteem van het universum (bijvoorbeeld een menselijke samenleving, de mensheid als geheel, of een individueel mens – welk systeem het extropianisme ook bestudeert) afneemt (wat vermoedelijk gelijk zou staan aan een toename van de "extropie"), dan betekent dit fysiek dat dit alleen mogelijk is ten koste van een overeenkomstige *grotere toename* van de entropie van het gehele systeem.

Aangezien transhumanisten beweren strikt rationeel en wetenschappelijk te denken, zouden ze moeten erkennen dat, volgens de tweede wet van de thermodynamica, de extropie van het gehele universum daarom altijd *moet afnemen*. Extropianisme is dus een *quasi-ethische* oriëntatie binnen het kader van de natuurwet, beperkt niet tot het gehele systeem, maar tot het eigen *subsysteem* en de superioriteit daarvan. Men zou kunnen spreken van een soort kosmisch darwinisme, waarin de "extropie" van een "systeem" het

criterium vormt voor superioriteit in een over het algemeen vijandig universum. Het vergroten ervan is alleen mogelijk ten koste van andere subsystemen. Zo ligt in de kern van het neologisme "extropianisme" de codering van wezenlijk *sociale* wrokgevoelens als een fysieke metafoor.

Singularitarianisme, een andere substroming van het transhumanisme (vertegenwoordigd door de letter S in de TESCREAL-bundel van Gebrus en Torres), hanteert eveneens een dergelijke strategie, die kan worden omschreven als wrokachtig fysicalisme. Centraal in deze stroming staat het idee van een technologische singulariteit, die op haar beurt nauw verbonden is met het begrip "intelligentie-explosie". Er is dan een punt in de toekomst waarna recursief zelfverbeterende AI-systemen zich in een onstuitbare, zelfversterkende dynamiek van "intelligentieverbetering" zullen bevinden.

De metafoor van de technologische singulariteit is gebaseerd op het concept van de astrofysische singulariteit uit de algemene relativiteitstheorie. Daar verwijst de term naar formeel afleidbare constellaties waarin de zwaartekracht, en daarmee de kromming van de ruimtetijd, geen gedefinieerde waarde aanneemt, maar, om het in de volksmond te zeggen, "oneindig groot" is. Dit zijn constellaties die niet als punten op de ruimtetijd kunnen worden afgebeeld, omdat de ruimtetijd daar "instort".

Dit betekent dat onder de omstandigheden van een zwaartekrachtsingulariteit de bekende natuurwetten, in het bijzonder, worden opgeschort. Dit is bijvoorbeeld het geval in het centrum van een zwart gat. Omdat singulariteiten zich buiten de causale waarnemingshorizon bevinden, kan er geen informatie uit het "binnenste" worden verkregen. Een singulariteit die naar buiten kan doordringen, is daarom ook onmogelijk. Daarom is het bestuderen van natuurwetten op deze punten eveneens onmogelijk.

Op vergelijkbare wijze stelt het singularitarisme in de context van AI dat het moment waarop een AI een bepaalde drempel van recursieve zelfverbetering overschrijdt (een "intelligentie-explosie") een fundamentele en onomkeerbare breuk in de menselijke geschiedenis zal betekenen. Het systeem zal dan zo intelligent worden dat wij, binnen het huidige kader en met onze beperkte huidige intelligentie, niet in staat zullen zijn om de omstandigheden ervan te classificeren, te voorspellen of zelfs te beoordelen. Vanuit het huidige perspectief zijn daarom geen zinvolle voorspellingen mogelijk over de biologische of sociale wetten, omstandigheden en structuren van deze toekomstige superintelligente levensvorm. —De perfecte ideologische basis om elke genuanceerde discussie over wenselijke of onwenselijke toekomsten met AI te smoren in een diffuse, quasi-messiaanse mist van transcendentie.

Dit idee bestaat al sinds het midden van de 20e eeuw. John von Neumann behandelde het onderwerp in verband met een hypothetische superintelligentie. Meer recent is de AI-singulariteit ingezet als een populair ideologisch concept dat een toekomst met en door AI vertegenwoordigt die we ons niet kunnen veroorloven te missen, maar die we noch kunnen beoordelen, bekritiseren, noch mede vormgeven.

92-95

Eugenetica 2.0

Waar eerdere stromingen binnen het transhumanisme zich voornamelijk richtten op individuele zelfoptimalisatie en de technologische verbetering van individuen, is de focus sinds de jaren 2000 steeds meer verschoven naar de technologische verbetering van *de mensheid als geheel* door middel van normatieve controle over haar evolutie.

Deze visie vindt haar onmiskenbare basis in de eugenetica, dat wil zeggen pseudowetenschappelijke en politieke programma's voor de vermeende verbetering van de genetische kwaliteit van de menselijke bevolking. Eugenetica was wijdverbreid in de VS (evenals in Duitsland en andere Centraal-Europese landen) aan het einde van de 19e en de eerste helft van de 20e eeuw en werd altijd beschouwd als een zogenaamd wetenschappelijk instrument om onmenselijke politieke maatregelen te legitimeren, zoals de onderdrukking van migratie of van mensen met ziekten en handicaps.

De term eugenetica werd in 1869 in Groot-Brittannië bedacht door Francis Galton, een neef van Charles Darwin. Eugenetica is gebaseerd op het idee dat de genetische diversiteit van een populatie (bijvoorbeeld het eigen land of "volk") strikt gecontroleerd moet worden door de voortplanting te reguleren, om zo de vermeende evolutionaire selectiedruk tussen volkeren te kunnen overleven. Eugenetische programma's beïnvloeden de menselijke voortplanting op een repressieve of positieve manier. Positieve benaderingen richten zich op het stimuleren van individuen met vermeende gewenste genetische eigenschappen om zich vaker voort te planten, terwijl repressieve programma's de voortplanting van individuen met vermeende ongewenste genen beperken of zelfs elimineren. In de 20e eeuw ging eugenetica dan ook vaak hand in hand met dwangmaatregelen van de staat, zoals gedwongen sterilisaties of huwelijksverboden, en ging zelfs zo ver dat de vervolging en moord op bepaalde groepen mensen die zogenaamd "inferieur of gebrekkig" genetisch materiaal bezaten of die werden geclassificeerd als "leven dat het leven niet waard is".

Moderne, door het transhumanisme geïnspireerde eugenetica presenteert zich daarentegen doorgaans als een geïndividualiseerde, op technologie gebaseerde vorm van zelfoptimalisatie. Voorstanders geloven dat vooruitgang in genetische manipulatie, reproductieve geneeskunde en AI-ondersteunde diagnostiek het binnenkort mogelijk zal maken om selectief specifieke genetische eigenschappen te kiezen of te modificeren – of het nu gaat om het voorkomen van erfelijke ziekten, het verbeteren van intelligentie of fysieke prestaties, of het verlengen van de levensduur. Deze ontwikkelingen roepen echter niet alleen ethische vragen op met betrekking tot sociale ongelijkheid en discriminatie, maar tonen ook een voortzetting van eugenetisch denken onder het mom van transhumanisme. Het idee dat mensen bewust hun eigen evolutie zouden moeten sturen en optimaliseren, is rechtstreeks afgeleid van de idealen van Julian Huxley, die niet alleen de grondlegger van het transhumanisme was, maar ook een prominent voorstander van eugenetica in de 20e eeuw.

Gebru en Torres waarschuwen dat veel van deze ideeën vandaag de dag nog steeds voortleven in de vorm van technologische ideologieën. In hun kritiek op de AGI- en techno-optimistische scene tonen ze aan dat ogenschijnlijk neutrale concepten zoals AGI of de "toekomst van de mensheid" vaak slechts bekende eugenetische ideeën verbergen. Ze spreken van een ideologische continuïteit waarin racistische, validistische en elitaire ideeën over 'wenselijke toekomsten' en 'leefbare levens' nieuw leven worden ingeblazen – zij het met verwijzing naar concepten als veiligheid, efficiëntie en mondiale optimalisatie. In transhumanistische kringen wordt deze historische connectie vaak gebagatelliseerd of afgedaan als achterhaald.[11]
Hoewel deze connectie in transhumanistische kringen vaak wordt gebagatelliseerd, blijft eugenetisch denken aantoonbaar voortbestaan in de context van AI- en technologie-ideologieën. Een indicatie hiervan is de fixatie op intelligentie als het centrale selectiecriterium: “
Zolang er geen 'superintelligentie' bestaat," luidt de aanname, "is het eerst nodig om steeds intelligentere mensen te ontwikkelen die vervolgens krachtigere AI ontwikkelen, totdat deze een cyclus van recursieve zelfverbetering kan ingaan en de menselijke intelligentie kan overtreffen." De kwantitatieve meting van intelligentie door middel van een "IQ-test" vindt echter zijn oorsprong in het werk van eugenetici (waaronder Alfred Binet, Henry Goddard en Lewis Terman) in het begin van de 20e eeuw en was nauw verbonden met politieke programma's voor sociale selectie.

Wanneer de hedendaagse tech-elite de verbetering van cognitieve prestaties opnieuw bestempelt als een cruciaal beschavingsprobleem en dit koppelt aan overwegingen over de gerichte manipulatie van de menselijke voortplanting, onthult dit een structurele continuïteit van eugenetische

denkpatronen. Hoe dit zich momenteel manifesteert in de concepten van selectief pronatalisme zal in het volgende hoofdstuk worden uitgelegd.

95-98

De rationalisten en hun existentieel-futuristische ethiek

Nadat het transhumanisme zich in de late 20e eeuw voornamelijk richtte op de wetenschappelijke verworvenheden van genetica en nanotechnologie, kreeg het in de 21e eeuw een impuls door de invloed van een nieuwe epistemologische stroming. De beweging van de zogenaamde *rationalisten* in de VS speelde hierin een centrale rol.[12]
De rationalisten kwamen voort uit een online gemeenschap in Californische tech- en wetenschappelijke kringen en moeten niet worden verward met de rationalistische school in de filosofie. Een van de centrale figuren van de rationalistische beweging is de Amerikaanse autodidact en zelfbenoemde AI-onderzoeker Eliezer Yudkowsky, die in 2009 het online platform *LessWrong* oprichtte. Daar verzamelde zich een los georganiseerde gemeenschap, aanvankelijk bezig met het toepassen van Bayesiaans denken op vraagstukken van rationele besluitvorming en cognitieve zelfverbetering, en uiteindelijk ook op technologische vraagstukken van de toekomst, met name in het licht van de potentiële ontwikkeling van bovenmenselijke AGI. In de jaren 2010 verwierf *LessWrong* aanzienlijke invloed in de tech- en startupwereld (zowel in Silicon Valley als internationaal), waar rationalistische netwerken samensmolten met de transhumanistische beweging. In 2012 werd het Center for Applied Rationality (CFAR) opgericht in Berkeley, Californië, door leden van de *LessWrong*-gemeenschap. Sindsdien organiseert CFAR workshops, coachingsessies en beurzenprogramma's.

Voor rationalisten biedt de Bayesiaanse waarschijnlijkheidstheorie (zie hoofdstuk 2) een benadering die erop gericht is om denken en oordelen in onzekere situaties te bevrijden van valse intuïties, cognitieve vertekeningen en onderbuikgevoelens. Vanuit hun perspectief is een dergelijke methodisch gecontroleerde manier van denken fundamenteel voor het omgaan met existentieel belangrijke vragen over toekomstige ontwikkelingen. Het zou gebruikt kunnen worden om de methodologie van transhumanistische discoursen over technologische toekomsten op een (vermeend) rationeel fundament te plaatsen: terwijl eerdere transhumanistische benaderingen zeer speculatief waren, gaven de rationalisten deze ideeën een formeel en logisch niveau, waardoor ze beter aansluiten bij academische en technologische contexten.

Volgens hun eigen enquêtes bestaat de rationalistische beweging voornamelijk uit mannen tussen de 20 en 35 jaar die afkomstig zijn uit samenlevingen in het mondiale Noorden en werkzaam zijn in de IT-sector of

aanverwante wetenschappelijke vakgebieden.[13] Vanwege de alomvattende claim om de "fouten van de menselijke rationaliteit" bloot te leggen en een zo "foutloos" mogelijk wereldbeeld te ontwikkelen, is de beweging bekritiseerd als elitair en arrogant.

Recentelijk zijn er onder haar aanhangers steeds meer antidemocratische tendensen waargenomen, bijvoorbeeld in verband met technocratische visies op de samenleving. Met name in de jaren 2010 trok *LessWrong* ook individuen aan uit de neoreactionaire beweging (NRX, zie hoofdstuk 5) die geïnteresseerd waren in eugenetica en evolutionaire psychologie en die de sociaal-darwinistische doctrines van het transhumanisme wilden formaliseren met behulp van rationalistische instrumenten. Hoewel Yudkowsky zich genoodzaakt voelde afstand te nemen van de neoreactionaire beweging[14] zijn er nog steeds overlappingen te zien tussen delen van de rationalistische gemeenschap en openlijk racistische en eugenetische stromingen, vooral in de context van de alt-rightbeweging en in de kringen rond Donald Trump.

99-101

Effectief Altruïsme

Het zogenaamde Effectief Altruïsme (EA, ook wel vertegenwoordigd door het acroniem TESCREAL van Gebrus en Torres) is een beweging die nauw verbonden is met het rationalisme, zowel qua ideeën als qua personeel en middelen, en die sinds de jaren 2010 ook aanzienlijke institutionele steun heeft verworven. EA kan worden omschreven als de toepassing van rationalistische epistemologische principes op moreel handelen: Morele beslissingen worden wiskundig gemodelleerd met betrekking tot hun onzekere of speculatieve gevolgen om "rationele" en "effectieve" beslissingen te nemen, zelfs in het licht van zeer hypothetische toekomstscenario's. Deze benadering wordt institutioneel ondersteund door organisaties uit de technologie- en academische sector.
EA kan worden omschreven als de toepassing van rationalistische epistemologische principes op moreel handelen: Morele beslissingen worden wiskundig gemodelleerd met betrekking tot hun onzekere of speculatieve gevolgen om "rationele" en "effectieve" beslissingen te nemen, zelfs in het licht van zeer hypothetische toekomstscenario's. De filosoof William MacAskill van de Universiteit van Oxford, een sleutelfiguur in de EA-beweging, definieerde het in 2017 als: "[E]ffectief altruïsme is het project om bewijs en rede te gebruiken om te bepalen hoe anderen zoveel mogelijk baat kunnen hebben, en om op basis daarvan te handelen."[15] Het basisidee van effectief altruïsme is het zoeken naar de meest efficiënte manieren om maximaal welzijn te bereiken voor het grootste aantal mensen en voelende wezens – vandaar de zelfbenaming als een vorm van altruïsme.

De beweging wordt niet alleen verenigd door de utilitaristisch geïnspireerde consensus *dat* men de gevolgen van iemands handelingen moet optimaliseren om zoveel mogelijk meetbaar goed te doen, maar ook door de bereidheid om de eigen (ideale of materiële) middelen in te zetten om dit doel te bereiken. Bovendien richt EA zich primair op het vinden van een rationalistische methodologie om te bepalen hoe het "goede" het meest effectief kan worden bereikt met de beschikbare middelen. Centraal hierin staat de toepassing van analytische instrumenten zoals kosten-batenanalyse, Bayesiaanse waarschijnlijkheidstheorie en de wiskundige modellering van morele dilemma's op vraagstukken als armoedebestrijding, gezondheidszorg of onderzoek naar AI-beveiliging. Wanneer voorstanders van EA spreken over "het maximaliseren van het goede", bedoelen ze altijd een numerieke hoeveelheid.

Als sociale beweging ontstond EA eind jaren 2000 in samenhang met de maatschappelijke en filantropische implementatie van rationalistische principes, die werden verspreid in de eerdergenoemde online gemeenschappen. "Filantropie" verwijst in deze context naar een vorm van liefdadigheid die wordt ondersteund door vermogende personen en techmiljardairs, ideologisch gericht op effectiviteit en wereldwijde impact. Een belangrijke rol in deze ontwikkeling werd gespeeld door de NGO GiveWell, opgericht in 2007 door hedgefondsmanagers Holden Karnofsky en Elie Hassenfeld. GiveWell evalueert en rangschikt hulporganisaties met behulp van financiële analysemethoden. De inmiddels bekendere organisatie Open Philanthropy is in 2014 voortgekomen uit dit netwerk. Tegelijkertijd werden organisaties zoals Giving What We Can (2009) en 80,000 Hours (2011) opgericht in de omgeving van de Universiteit van Oxford. Deze twee groepen fuseerden in 2012 tot het Centre for Effective Altruism.

Zo is effectief altruïsme uitgegroeid tot een sterk geïnstitutionaliseerde en goed gefinancierde beweging. Via een netwerk van denktanks, filantropische organisaties, ngo's, stichtingen, startups en academische instellingen beïnvloedt deze beweging actief politieke beslissingen, onderzoeksprioriteiten en economische investeringen. Er bestaan met name nauwe banden – qua personeel, ideeën en financiën – met de techindustrie, vooral Silicon Valley, waar veel van de grootste supporters van effectief altruïsme gevestigd zijn, waaronder Dustin Moskovitz (Facebook, Open Philanthropy), Jaan Tallinn (Skype), Vitalik Buterin (Ethereum) en Sam Bankman-Fried (FTX, voordat zijn financiële imperium instortte). Oxford, met de filosofen Toby Ord en William McAskill, werd lange tijd beschouwd als een academisch centrum van de EA-beweging. Lokale EA-groepen zijn echter aan talloze universiteiten wereldwijd opgericht – waaronder Duitse universiteiten, waar ze vaak functioneren als studentenverenigingen die jonge studenten proberen te werven als aanhangers, door hen te lokken met

genereuze beurzen van EA-netwerkorganisaties. Men kan deze lobby-inspanning van Effectief Altruïsme zien als een opstapje naar de wereld van elitaire tech-ideologieën.[16]

Hoewel effectief altruïsme (EA) in eerste instantie overtuigend lijkt, heeft het te maken gehad met serieuze inhoudelijke en institutionele kritiek:

Een belangrijk probleem met EA is het consequentialistische karakter ervan, oftewel de neiging om morele beslissingen te beoordelen op basis van de verwachte gevolgen. Hoewel McAskill en Ord EA scherp proberen te onderscheiden van een naïef, muggenzifterig utilitarisme, blijft het gedreven door de impuls om beslissingen – bijvoorbeeld bij het evalueren van carrièreopties of door donaties gefinancierde organisaties – te baseren op een wiskundige impactanalyse, en dus op een numerieke vergelijking van verwachte gevolgen en effecten.

De simplistische, reductionistische, technocratische en vaak arrogante benadering van morele vraagstukken blijft problematisch. EA beschouwt ethische beslissingen primair als wiskundige optimalisatieproblemen en negeert systematisch sociale onderhandelingsprocessen en politieke dimensies in de strijd tegen ongelijkheid. Technologie is altijd de oplossing, niet onderdeel van het probleem. Het bestaande systeem wordt nooit in twijfel getrokken; Structurele problemen zoals kolonialisme, kapitalisme of machtsaccumulatie worden als te complex of te moeilijk te kwantificeren beschouwd en worden daarom genegeerd ten gunste van kortetermijnoplossingen die op schaal kunnen worden toegepast. In EA-strategieën wordt de verantwoordelijkheid voor maatschappelijke problemen verschoven van publieke instellingen naar private actoren – een ontwikkeling die doet denken aan de technocratische herinterpretatie van staatsfuncties, zoals besproken in hoofdstuk 2 aan de hand van het voorbeeld van openbaar bestuur. Institutioneel gezien fungeert effectief altruïsme in zijn huidige vorm uiteindelijk als de ethische arm van het Silicon Valley-kapitalisme. Met enorme financiële middelen verspreidt de beweging kapitalistische denkwijzen in de burgermaatschappij. Daar worden maatschappelijke problemen geformuleerd als technologische uitdagingen die kunnen worden opgelost met algoritmen en probabilistische logica.

101-105

Langetermijnisme

Bijzonder relevant in de context van reactionaire technologie-ideologieën is de vernauwing van de focus binnen de EA-beweging in het afgelopen decennium. Deze focus is steeds meer verschoven van acute mondiale problemen zoals armoede of gezondheidszorg naar speculatieve scenario's

in de verre toekomst – met name in de context van AI. De centrale focus van deze stroming ligt op de beschavingsrisico's die een hypothetische, vijandige AGI over miljoenen jaren of langer met zich meebrengt (zie hoofdstuk 3).

In deze context verwijst de term "*Langetermijnisme*" naar de ethische doctrine dat risico's op de ultralange termijn *moreel moeten worden overwogen* of zelfs prioriteit moeten krijgen. De term 'ultralange termijn' verwijst minder naar de volgende generaties, wier levensonderhoud aanzienlijk zal worden beïnvloed door klimaatverandering, maar eerder naar mensen die over miljoenen, miljarden of triljoenen jaren zullen leven.[17]

De term 'Langetermijnisme' werd eind jaren 2010 bedacht door de twee Oxfordse filosofen MacAskill en Ord.[18] Hun gedachtegoed combineert effectieve altruïstische ethiek met de technofuturistische visies van transhumanisme, zoals bepleit door Nick Bostrom. Een voorbeeldige redenering volgt grofweg drie stappen:[19]

1. Door de ontwikkeling van superintelligentie zou de mensheid – als de superintelligentie welwillend is – zich in de ultraverre toekomst als digitale mensen over de kosmos kunnen verspreiden. Ze zou dan een enorme populatie vormen, die de huidige menselijke populatie met meerdere ordes van grootte zou overtreffen.

2. Als de ontwikkeling van AGI echter zou leiden tot een vijandige superintelligentie, of als de ontwikkeling ervan zou worden verhinderd of vertraagd, zou het potentiële voortbestaan van deze kosmische menselijke populatie in de verre toekomst in gevaar komen. Bovendien zou de ontwikkeling van een kwaadaardige superintelligentie het bestaande leven op aarde ernstig in gevaar kunnen brengen en mogelijk zelfs kunnen uitroeien.

3. Aangezien we vandaag de dag nog steeds invloed hebben op de ontwikkeling van AGI (zowel wat betreft de ontwikkeling ervan als de vraag of deze welwillend zal zijn), moeten we moreel gezien het welzijn van de menselijke populatie in de verre toekomst in overweging nemen, omdat, gezien de enorme omvang van deze potentiële populatie, een groot aantal "gelukkige mensenlevens" ervan afhangt.[20]

Dit vereist enige uitleg. Om de bedreigende absurditeit van dit standpunt te begrijpen, is het de moeite waard om kort de aanvankelijke premisse van explosieve menselijke groei in de verre toekomst te onderzoeken. Dit lijkt wellicht al twijfelachtig in het licht van de dreiging van een klimaatcatastrofe op middellange termijn, die door wetenschappers als zeer waarschijnlijk wordt beschouwd. Wetenschappers die zich op de lange termijn richten,

houden zich echter niet alleen bezig met de huidige aardse bevolking van ongeveer 10^{12}, maar gaan er, in overeenstemming met de techno-utopische visie van een kosmische evolutie van de mensheid, van uit dat mensen de ruimte zullen koloniseren met behulp van AI. Hier komt het transhumanistische geloof in het overwinnen van de biologische beperkingen van de mensheid door de ontwikkeling van superintelligentie in beeld. Volgens dit geloof zal de menselijke soort, door de digitalisering van geest en bewustzijn, haar aardse beperkingen overwinnen en zich min of meer "met de snelheid van het licht" in de ruimte verspreiden.[21] Digitale mensen zouden dan uiteindelijk bestaan in gigantische, interstellare computersimulaties.[22] Nick Bostrom, een van de belangrijkste denkers, berekende in een essay uit 2003 het aantal digitale mensen dat alleen al in de Virgo-cluster zou kunnen leven, dat wil zeggen in onze directe galactische omgeving. Hij baseert een 'conservatieve schatting' van dit aantal op een 'schatting van de rekenkracht die uit een ster kan worden gehaald', vermenigvuldigd met het aantal sterren in de Virgo-cluster en gedeeld door de rekenintensiteit van een hersensimulatie gedurende een mensenleven. Hieruit concludeert hij 'dat er in elke eeuw dat de kolonisatie van onze lokale supercluster wordt uitgesteld, potentieel ongeveer 10^{38} mensenlevens verloren gaan; ofwel, equivalent, ongeveer 10^{29} potentiële mensenlevens per seconde' (ibid.). Bostrom benadrukt dat het niet de exacte waarde van het getal is die telt, maar de enorme omvang ervan – als een soort ethische maatstaf. Vanuit een utilitaristisch perspectief zou dit enorme potentiële verlies aan mensenlevens een enorm waardeverlies betekenen. 'Weinig andere filantropische initiatieven zouden kunnen hopen een dergelijk utilitaristisch rendement te behalen.'[23]
Deze speculatieve ideeën met hun grove ethische conclusies zouden zeker niet zoveel aandacht krijgen als de politiek en economisch invloedrijke miljardairs van vandaag niet alles in het werk zouden stellen om deze doelen te bereiken. Zo heeft Elon Musks bedrijf SpaceX aangekondigd dat het binnen enkele jaren naar Mars wil reizen; Musks bedrijf Neuralink streeft ernaar een brein-computerinterface te ontwikkelen. Binnen de kringen van de technologische elite heeft het idee van ruimtekolonisatie, in combinatie met de ethiek van het langetermijndenken – volgens welke het heden ethisch onbeduidend is in vergelijking met de ultraverre toekomst – zich gevestigd als een centraal referentiekader. Het krijgt steeds meer invloed op de toewijzing van investeringen, onderzoekscontracten en politieke besluitvorming.

Een groot deel van de leden van de EA-beweging houdt zich nu bezig met het Langetermijnisme. Politieke besluitvormers en internationale organisaties grijpen ook steeds vaker terug op argumenten die zich richten op de lange termijn, zoals blijkt uit voorbeelden als de "AI Safety Summit" in Bletchley Park (VK, 2023) en de "Artificial Intelligence Action Summit" (Parijs, 2025). Het enorme lobbysucces van het Langetermijnisme leidt er niet alleen toe

dat het onmiskenbare risico van klimaatverandering wordt gebagatelliseerd, maar ook dat de manifeste maatschappelijke gevolgen van AI in het heden – uitbuiting, sociale ongelijkheid, discriminatie en machtsconcentratie – als ondergeschikt worden beschouwd aan de apocalyptische kansen en risico's van AI op de ultralange termijn.

105-109

Seculiere eschatologie

In haar essay "*Ghost in the Cloud*" uit 2017 probeert de Amerikaanse schrijfster Meghan O'Gieblyn uit te leggen waarom zulke vergezochte, hypothetische denkbeelden als de extreme vormen van transhumanisme überhaupt in de samenleving kunnen worden gehandhaafd en verspreid. Dit biedt een interessante interpretatie van de wisselwerking tussen de buitenste laag van het publieke discours en de onderliggende lagen, waaronder de discoursen van actoren in de techwereld. O'Gieblyn beschrijft hoe de techno-utopische verhalen die herhaaldelijk worden aangehaald in transhumanistische ideologieën en populaire AI-verhalen onbewust een oud religieus motief aanboren: ze herformuleren de belofte van eeuwig leven – een komende opstanding – in een staat van verlossing, waarin alle aardse beperkingen en lijden zullen worden overwonnen.[24] Klassieke noties van onsterfelijkheid, zoals die welke de christelijke tradities hebben gevormd, duiken opnieuw op in geseculariseerde vorm – bijvoorbeeld als het idee om het menselijk bewustzijn naar de cloud te uploaden of meester te worden van de menselijke evolutie door de ontwikkeling van AI. Deze verhalen oefenen een vergelijkbare emotionele kracht uit op hun aanhangers als het bijbelse verhaal. Het idee van transcendentie, alleen neemt technologie hier de plaats in van het goddelijke.

Uitgaande van de premisse dat deze belofte van een toekomstige staat van verlossing en redding grotendeels onbetwist blijft, spelen technologische ideologieën zoals Langetermijnismeen transhumanisme in op mystieke verlangens zonder hun werkelijke religieuze wortels te hoeven erkennen. O'Gieblyn legt uit dat het concept van verlossing dat deze technologische ideologieën bieden zo verleidelijk is "omdat het belooft, door middel van wetenschappelijke methoden, de transcendente hoop te herstellen die de wetenschap zelf heeft uitgedoofd", zonder dat deze verlossing zich hoeft te buigen over "de netelige problemen van goddelijke gerechtigheid" op de Dag des Oordeels. De ethiek die hieruit voortvloeit, betreft daarom niet het vrome leven van het individu dat uiteindelijk naar de hemel of de hel gaat, maar eerder de succesvolle technologische verbetering van de beschaving als geheel (inclusief de overwinning in de internationale wapenwedloop, die zal bepalen of de "goede jongens" of de "slechte jongens" AGI ontwikkelen). Dit verschuift de focus van verlossing van individuele moraliteit naar het

collectieve lot van een beschaving, waarvan het voortbestaan en de triomf worden gepresenteerd als een quasi-transcendente waarde. Volgens O'Gieblyn vertegenwoordigt dit een "evolutionaire benadering voor eschatologen – dat wil zeggen, voor de christelijke leer van de laatste dingen, zoals het einde van de wereld, het Laatste Oordeel en het eeuwige leven."[25]
Deze structuur – de belofte van verlossing door radicale transformatie, de interpretatie van het heden als een crisis en de toekomst als een zuiverende wedergeboorte – komt in haar formele logica overeen met centrale patronen van fascistische ideologieën. Ook in deze ideologieën verschijnt de geschiedenis als een laatste strijd die culmineert in een collectieve wedergeboorte.

Het ideologische kader van AI-technologie is dus niet alleen religieus geladen, maar blijkt ook compatibel met een sacralisering van het politieke: het neemt motieven over zoals de "nationale wedergeboorte" of het "Duizendjarige Rijk", die kenmerkend zijn voor fascistische ideologieën.

De technologische singulariteit neemt structureel de rol aan van een seculier Laatste Oordeel: ze belooft zuivering door de vernietiging van het oude, het overwinnen van zwakte en de creatie van een hoger, postmenselijk collectief subject. De ideologische verwantschap ligt niet in identieke inhoud, maar in de politieke functie van dergelijke verlossingsverhalen: ze legitimeren radicale breuken en onderwerpen het heden aan een toekomstig doel dat als historisch noodzakelijk wordt beschouwd.

109-111

Politiek in het hier en nu

Met name langetermijnideeën vormen de mystieke basis van sommige denkwijzen die circuleren in het publieke AI-debat. Deze ideeën vormen samen een giftige mix van een radicaal zelfingenomen en sociaal-darwinistisch technologiebeleid. Deze verhalen stellen actoren in de techindustrie in staat zichzelf af te schilderen als moreel verantwoordelijke visionairs die zich zorgen maken over het voortbestaan van de mensheid op de lange termijn.

Deze pseudo-ethische benadering *negeert strategisch kwesties van macht en verdeling:* zo wordt bijvoorbeeld voorbijgegaan aan de vraag welk deel van de mensheid vandaag de dag daadwerkelijk in staat zou zijn om andere planeten te koloniseren, hun geest naar de cloud te uploaden of zelfs te delen in de rijkdom die de AI-industrie vandaag de dag vergaart.

De technologie-industrie is een economische sector die een enorme planetaire factor is geworden in de exploitatie van menselijke en materiële hulpbronnen en de ecologische bedreiging die zij vormt, terwijl zij tegelijkertijd nieuwe dimensies van persoonlijke vermogensopbouw genereert.[26] Vanuit een langetermijnperspectief lijken de huidige levenskwaliteit en algemene welvaart, sociale rechtvaardigheid, pluralisme, democratie en ecologische duurzaamheid echter, vergeleken met het potentiële welzijn van een enorme menselijke bevolking in de ultraverre toekomst, slechts variabelen in een utilitaristische kosten-batenanalyse, waarbij de astronomische toekomst altijd voorrang krijgt boven het heden.

Het is daarom cruciaal om dergelijke toekomstverhalen te begrijpen als *politieke* interventies die specifieke belangen *in het heden* willen afdwingen. Dit leidt de aandacht af van de werkelijke onrechtvaardigheden en crises die niet alleen *massaal worden bevorderd door de AI-industrie, maar ook winstgevend worden uitgebuit.* Langetermijnargumenten legitimeren de negatieve gevolgen van AI vandaag de dag, waaronder met name: economische en ecologische uitbuiting, sociale ongelijkheid en belastingontduiking, als een offer voor een zogenaamd hoger moreel goed in de verre toekomst.

Het volgende hoofdstuk zal aantonen dat niet alleen de huidige sociaaleconomische problemen worden afgezet tegen utopische technologische toekomstvisies, maar dat de opkomst van autoritaire machtsstructuren en neofascistische regimes ook wordt gelegitimeerd en in sommige gevallen zelfs begeerd door de superioriteit van ongebreidelde technologische vooruitgang en een ultieme belofte van verlossing in de verre toekomst. Sommige van deze technologische ideologieën zijn ronduit antidemocratisch, sociaal-darwinistisch en neoreactionair. Ze zijn nu samengesmolten tot een bredere politieke stroming die niet langer beperkt is tot techkringen en hun filantropische organisaties, maar samen met de alt-right en de nieuwe rechtervleugel is uitgegroeid tot een grote en bedreigende, fascistische politieke beweging.

111-113