**OVERIG MATERIAAL TRANSPARANTIE & AI   
(Crash course Five)**

Wij bieden crashcourses aan op [www.tict.io](http://www.tict.io). Deze zijn in het Engels. Daarnaast hebben we in het verleden veel materiaal gemaakt in het Nederlands, dit is allemaal verzameld in dit word-document.

In dit document staat alles rond crash course vier, Big Data. Het is verzameld (oud) materiaal, dus we raden je vooral aan om lekker te grasduinen!

De Engelse geupdate online course staat hier: [www.tict.io/course-five](http://www.tict.io/course-five).

**Introductie**  
Op technofilosofie.com doen we aan **toegepaste technofilosofie.**Het is ons doel om zoveel mogelijk mensen na te laten denken over de impact van technologie. Wij denken namelijk dat iedereen dat zou moeten doen. Iedereen zou een beetje een technofilosoof moeten zijn. Of je nu technologie ontwerpt, programmeert, implementeert of gewoon gebruikt. Of je nu een professional bent, of een vader, een coach, een vriend, vakantieganger of gewoon een mens.

Moderne digitale technologie rukt op en wij moeten leren om ons daartoe te verhouden. Het gaat daarbij niet om een abstract gegeven, maar gewoon om onze relatie alledaagse, hedendaagse technologie. Smartphones, e-mail, platformen zoals Uber, Tinder, sociale media, sensoren, slimme meters, en ga zo maar door.

**DAT IS ONZE MISSIE: ZOVEEL MOGELIJK MENSEN ZO GOED MOGELIJK TE LATEN NADENKEN OVER DE IMPACT VAN TECHNOLOGIE**.

We bieden daarom ook materiaal aan, zoals in dit word-document. Hierin staan verwijzingen naar artikelen, video's e.d. De tekst valt onder creative commons, maar de afbeeldingen, verwijzingen, clips en video's niet altijd. Het is dus belangrijk dat je je daar zelf van op de hoogte stelt, als je materiaal wilt gebruiken. Bij twijfel, check de [website](https://creativecommons.nl/) van creative commons.

Op deze pagina staat een **long read** over kunstmatige intelligentie. De long read gaat over voordelen en nadelen. Over ethische vragen en over al die zaken waar je over na moet denken als je iets gaat doen met kunstmatige intelligentie of ermee te maken krijgt. In de long read staan verwijzingen naar artikelen, video's e.d. De long read maar ook alle onderliggende pagina's en eigenlijk dit hele platform valt onder Creative Commons, maar de afbeeldingen, verwijzingen, clips en video's niet altijd. Het is dus belangrijk dat je je  daar zelf van op de hoogte stelt, als je materiaal wilt gebruiken. Bij twijfel, check de [website](https://creativecommons.nl/) van creative commons.

[Hier vind je de PowerPoint](https://maken.wikiwijs.nl/136737/Technofilosofie___Aandacht#!page-5633958) van de masterclass.

De volgende vragen gaan we proberen te beantwoorden:

1. Wat is kunstmatige intelligentie. Wat zijn verwachte ontwikkelingen op lange termijn? Worden computers slimmer dan mensen? Are we fucked?
2. Hoe zit het met automatische beslissystemen?
3. Zomaar wat losse opmerkingen over andere AI - ontwikkelingen!
4. Moeten we AI reguleren en hoe doen we dat dan?
5. Krijgen machines verbeeldingskracht?
6. Hebben we straks nog banen?

Bijlage A: Extra Achtergrond informatie  
Bijlage B: Genieten van AI

**(1) Wat is kunstmatige intelligentie? Wat zijn verwachte ontwikkelingen en worden computers slimmer dan mensen? Are we fucked?**Het is best ingewikkeld om deze vraag te beantwoorden. Maar het is wel belangrijk. Je zou maar in discussies terecht komen met andere mensen over kunstmatige intelligentie, maar allebei over volstrekt iets anders praten. Dat is weinig intelligent, ben ik bang. Begrippen als kunstmatig, leven, intelligentie en bewustzijn kunnen veel misverstanden veroorzaken. Ik heb niet de illusie dat wij op technofilosofie.com wél de wijsheid in pacht hebben. Echter, wat we wel kunnen doen, is een aantal definities geven welke we vervolgens consequent op deze site zullen hanteren. Daarvoor maken we dankbaar gebruik van het werk van [Max Tegmark](https://www.bol.com/nl/p/life-3-0/9200000078814498) (link naar boek).

De eerste vraag is, **wat is leven?** Laten we zeggen dat een leven een proces is, dat zijn complexiteit kan handhaven en zich kan repliceren. Als je er zo naar kijkt, heb je 3 verschillende stadiums van leven. Life 1.0 is simpel. Zeg, een worm. Het kan overleven en het kan zich repliceren. Life 2.0 is complexer. Het kan hetzelfde als life 1.0 maar het kan ook zijn software ontwerpen. Bijvoorbeeld, iemand die iets leert of afleert verbetert zijn software. Life 3.0 is technologisch. Dit leven kan overleven, repliceren, zijn software ontwerpen en zijn hardware ontwerpen. Momenteel zitten we in de fase van life 2.0 met componenten van life 3.0. Wij mensen kunnen bijvoorbeeld steeds beter onze eigen hardware ontwerpen. Denk maar aan een bril, een prothese, een pace-maker en ga zo maar door. Je zou al fundamentele vragen kunnen stellen of een slim ontworpen computervirus volgens deze definitie leven is? Kan die overleven? Zich repliceren? Zich aanpassen? Binnen bepaalde kaders waarschijnlijk wel.

De tweede vraag die je eigenlijk moet beantwoorden is: **wat is intelligentie?** Dat is best wel een ingewikkelde vraag, waar veel debat over kan zijn. De beste manier om dat debat te omzeilen, is opnieuw om zelf te kiezen voor een definitie en die consequent te hanteren in het vervolg. Op technofilosofie.com hanteren we de definitie: **intelligentie is het vermogen om complexe doelen te bereiken**. Als je die definitie hanteert heb je verschillende verdere verfijningen van deze definitie. Je hebt bijvoorbeeld**beperkte intelligentie.** Dat is het vermogen om een beperkt aantal doelstellingen te bereiken. Dat is een calculator, een schaakcomputer, een zelfrijdende auto, een computerdokter. Het kunnen dus hele intelligente dingen zijn die er gebeuren, maar wel beperkt. Daarnaast heb je **algemene intelligentie**. Dat is het vermogen om nagenoeg iedere doelstelling te bereiken, waaronder het aanleren van nieuwe vermogens.

**Kunstmatig** laat zich dan het meest eenvoudig omschrijven door: niet biologisch. Als je die twee combineert krijg je kunstmatige intelligentie. In de literatuur wordt vaak gesproken over **Algemene Kunstmatige Intelligentie (AKI).** Dat is het vermogen om alle cognitieve taken (taken waar je bij na moet denken) net zo goed uit te voeren als de mens. Daarnaast heb je nog de term superintelligentie. Dat is AKI die ver uitstijgt boven het menselijk niveau. Zo'n **superintelligentie** is het gevolg van AKI die zich snel zelf verbeterd en dat leidt tot een **intelligentie explosie**. Een ander woord daarvoor is **singulariteit**.

Klinkt best goed toch? De wereld kan wel een beetje meer intelligentie gebruiken. Er is al genoeg domheid in de wereld! Of ligt het toch wat genuanceerder? Het begint in ieder geval met de vraag: hoe snel gaat dit allemaal? En wanneer hebben we dan zo'n AKI of superintelligentie?

Als je bovenstaande leest, dan is kunstmatige intelligentie niet zo heel bijzonder. Dat hebben we namelijk al lang. Een rekenmachine is kunstmatige intelligentie. De schaakcomputer die Garry Kasparov versloeg in 1997 (Deep Blue) is kunstmatige intelligentie. Toch zal iemand niet snel zijn rekenmachine zien als kunstmatige intelligentie. Dat komt waarschijnlijk omdat een rekenmachine niet kan **leren**. De schaakcomputer uit 1997 voert slim volgende zetten uit, maar kan niet leren van zijn fouten. De belangrijke ontwikkeling is daarom dat we uitgevogeld hebben hoe we kunnen programmeren met behulp van **neurale netwerken**. We hebben geleerd hoe we machines kunnen programmeren, die leren.

Wil je weten, hoe dat werkt, dan is dit een video met een goede uitleg.

<https://www.youtube.com/watch?v=3bJ7RChxMWQ>

Hoe het allemaal precies werkt, is voor een technofilosoof niet heel belangrijk, maar het idee is dat de computers vooruit kunnen denken. Dat er een bepaalde output kan worden gedefinieerd en dat als er dan voldoende input is (trainingsdata) de computer begint te leren van zijn fouten en steeds beter wordt.

Om een indruk te krijgen van hoe indrukwekkend dit is, het volgende voorbeeld uit de schaakwereld. In 1997 won Deep Blue, de schaakcomputer van IBM van schaakgrootmeester Garry Kasparov. Maar dat was ook het enige wat Deep Blue kon. Je had er moeiteloos van kunnen winnen met boter-kaas-en-eieren. Daarna ontstond een korte periode waarin kunstmatige en intelligentie samenwerkten en tegen elkaar speelden. We noemden die teams Centaurs. Het wordt vaak als voorbeeld gebruikt van hoe robot en mens kunnen samenwerken. Iemand die zoiets verteld, gebruikt waarschijnlijk ook de term **Cobot**. Echter, wat er vaak niet bij verteld wordt, is dat inmiddels Centaurs bijna niet meer bestaan, omdat de computers veel te goed zijn en de mensen dus irrelevant. Dus speelden de schaakcomputers tegen elkaar en spelen mensen vrijwel geen rol meer (een beetje net als Formule 1 - of niet?). De beste schaakcomputer van allemaal was Stockfish 8. En die was nog klassiek geprogrammeerd door mensen. Het was net als Deep Blue vooral een computer die heel veel rekenkracht had, en dus heel snel, heel veel scenario's kon uitrekenen en daar de beste zet uit kon kiezen.

Maar onlangs speelde deze computer tegen een systeem dat neuraal geprogrammeerd was, namelijk AlphaZero. Dit systeem had nog nooit geschaakt, maar kende de output, namelijk: winnen. Dus oefende het systeem door miljoenen potjes tegen zichzelf te spelen en binnen 4 uur wist het 100 van de 28 partijen te winnen van Stockfish 8.

Als je even Googlet (ook kunstmatige intelligentie) krijg je vele indrukwekkende voorbeelden van zelflerende systemen. Omdat computercapaciteit toeneemt, omdat data-opslag steeds goedkoper is én omdat er steeds meer data is, kunnen deze systemen steeds beter getraind worden en leren ze steeds sneller. Eén keer in de 2 jaar komt een groep kunstmatig intelligentie onderzoekers bij elkaar en dan wordt de vraag gesteld: wanneer denken jullie dat er de eerste universele kunstmatige intelligentie zal zijn. Dat getal wordt dan gemiddeld en de laatste keer was het gemiddelde 2047. Best snel.

Wat betekent dit nu? Worden computers straks echt slimmer dan mensen? En hebben we dan een groot probleem? Wat doen we dan? Of krijgen we computers die heel erg doelgericht zijn? Een AI die niet van je houdt, die je ook niet haat, maar die weet dat je gemaakt bent van atomen die je kunt gebruiken voor iets anders. Bijvoorbeeld paperclips. De filosoof Nick Bostrom deed een [gedachtenexperiment](https://www.theverge.com/tldr/2017/10/11/16457742/ai-paperclips-thought-experiment-game-frank-lantz) (link naar paper) waarbij een AI de opdracht krijgt om paperclips te maken, maar daar zo fanatiek in is (of verkeerd geÃ¯nstrueerd) dat hij de hele wereld omzet in paperclips en vervolgens begint met de verpaperclipisering (is dat een woord?) van het heelal. De parabel van Bostrom is zo populair dat er een [game](https://nickbostrom.com/ethics/ai.html) van is gemaakt (link naar artikel over game en game).

In zijn boek Life 3.0 schetst Max Tegmark verschillende toekomstscenario's die hij met zijn groep van kunstmatige intelligentie specialisten heeft geschetst. We gaan even kort door deze scenario's heen, en dan kijken we naar de stappen die we nu kunnen zetten. In de afbeelding hieronder, zie je de mogelijke scenario's. We beginnen met het meest waarschijnlijke scenario (volgens de KI - onderzoekers, maar misschien doen die ook aan wensdenken). Vergeet niet, ze voorzien dit soort scenario's vanaf 2047. Het klinkt soms allemaal een beetje vergezocht, maar dat is voor de KI-onderzoekers zeker niet het geval.

Lees de scenario's er gerust een keer op na ([link naar site](https://futureoflife.org/superintelligence-survey/?cn-reloaded=1)). Ze lopen uiteen van een utopie tot een wereld waarin wij gehouden worden in een dierentuin.

En dan zijn er nog mensen die geen idee hebben welk scenario er gaat komen én er zijn mensen die alle scenario's niet verwelkomen. En natuurlijk zijn er ook onderzoekers die helemaal niet geloven in een universele kunstmatige intelligentie, dus voor wie deze scenario's complete luchtfietserij zijn. Je zorgen maken over een universele kunstmatige intelligentie staat gelijk aan je zorgen maken over overbevolking op Mars, stellen zij.

Die laatste groep is ook interessant. Want er zijn best een aantal argumenten aan te voeren die er op wijzen dat de ontwikkeling ook wel eens tegen kan vallen. Die zetten we hier op een rij.

A. De AI - Winter  
In de geschiedenis hebben we een aantal periodes gehad van héél véél optimisme rond kunstmatige intelligentie. Alan Turing schreef al in 1950 het provocerende essay: Can Machines Think? en in de jaren daarna werd er zwaar geïnvesteerd in kunstmatige intelligentie, zeker op het gebied van vertalen. Uiteraard naar en van het Russisch. Koude oorlog, weet je nog? Echter, daarna volgde een periode van minder enthousiasme. Dat noemen we een [**AI - winter**](https://www.popsci.com/ai-winter-artificial-intelligence#page-2) (link naar Artikel op Popsi). Ook nu, door doorbraken die we hierboven hebben beschreven, zoals meer computerkracht en nieuwe (neurale) manieren van programmeren zitten we midden in een hete zomer, maar in de verte zien steeds meer mensen een nieuwe winter komen.

Een mooi voorbeeld is de [zelfrijdende auto](https://www.theverge.com/2018/7/3/17530232/self-driving-ai-winter-full-autonomy-waymo-tesla-uber) (link naar artikel op The Verge). Die had er volgens Elon Musk al in 2018 moeten zijn. Helemaal autonoom, maar dat gaat niet gebeuren en het zou ook nog wel eens een generatie kunnen duren. Dat komt, voornamelijk door een proces genaamd, generalisatie. Simpel gezegd, door het miljoenen plaatjes te laten zien, kun je een computer perfect 'trainen' om een tijger te herkennen. Hetzelfde kun je doen voor een poes. We noemen dat interpolatie. Maar, als je dan vraagt, en nu een poema herkennen, die zit er ergens tussen in, daar snapt een computer niets van. Een kind van 4 wel. En de weg zit vol 'poema's'.

Hieronder een mooi voorbeeld dat laat zien hoe complex het nog is en hier nog een [grondig artikel](https://blog.piekniewski.info/2018/05/28/ai-winter-is-well-on-its-way/) (van de piekniewski blog).

B. Machines die denken? Echt waar?  
In de geschiedenis zijn de hersenen van de mensen vaak vergeleken met de op dat moment heersende 'technologie'.  Eerst dachten we dat er een klein mensje in ons hoofd zat (maar ja, wat zat er dan in het hoofd van dat kleine mensje). Daarna dachten we dat er raderen in rond draaide, of dat het meer een geavanceerd systeem was voor telegraaf - communicatie en nu denken we dat het een computer is. Maar snappen we wel hoe onze hersens werken? Ik hoefde mijn dochter van vier geen miljoenen plaatjes van katten laten zien, alvorens ze een kat kon herkennen. Godzijdank. En als we de hersenen niet snappen, snappen we dan wel hoe we computers bouwen, die slimmer zijn dan mensen?

C. Exponentiële onzinKalkoenen zijn sinds 1929 verdubbeld in grootte. Dat betekent, als deze trend zich doorzet, dat ze over 150 jaar net zo groot zullen zijn als mensen en over 6000 jaar net zo groot als onze planeet. Onzin, zegt u. Inderdaad, groei uit het verleden is geen garantie voor de toekomst. Groeit wijzigt, vlakt af, versnelt, stopt. Dat geldt ook voor de groei van computerkracht. Er is geen reden om aan te nemen dat de Wet van Moore zich door zal zetten, omdat het in het verleden zo was.

D. Het einde van de boom  
Wanneer je steeds sneller in staat bent, als rups bijvoorbeeld, om een boom te beklimmen, dan kom je op een gegeven moment (denk nog eens aan exponentiële groei) heel snel boven in de boom. Het betekent echter nog niet dat je dan op weg bent naar de maan. Je bent gewoon boven in de boom.

Overigens betekenen al deze zaken dat computers misschien wel nooit slimmer worden dan mensen. Het betekent niet dat kunstmatige intelligentie geen enorme impact heeft. Daar gaat de volgende paragraaf over. Én voor het geval dat de mensen van het Future of Life instituut toch gelijk hebben, en we wél algemene kunstmatige intelligentie hebben in 2047, is het fijn dat ze alvast de [Asilomar Principles](https://futureoflife.org/ai-principles/), ethische principes voor AI, hebben uitgewerkt. Je weet immers nooit....

**(2) Alles over automatische beslissystemen**

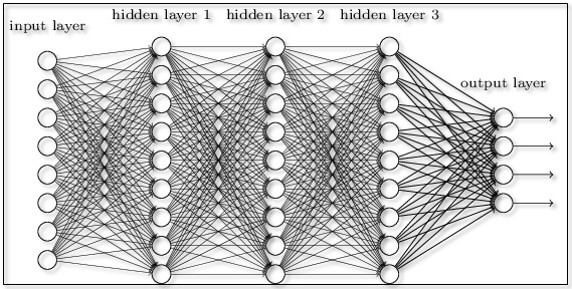
Oké, niet alles, maar wel een paar dingen.

Maar voordat we praten over deze systemen is het goed om te constateren dat overal om ons heen al geavanceerde vormen van kunstmatige intelligentie zijn (machine learning, deep learning).  Denk aan bijna alledaagse toepassingen als Google (Maps), tot de newsfeed van Facebook, de trending topics op Twitter en de bewakingscamera's op schiphol.

Een mooi voorbeeld is Tinder. Die gebruikt - onder meer - een techniek genaamd clustering. Tinder werkt namelijk met de elo-score. Dat is een term die afkomstig is uit de schaakwereld. Een beetje nerdy, maar Tinder is dan ook niet bedacht door mensen die het goed deden in de kroeg. Als je iemand swipet met een hogere score, en die swipet je terug, dan gaat je score omhoog en vice versa. Op die manier word je langzaam ingedeeld bij mensen bij wie de kans op een match het grootste is. Het betekent echter ook dat als je alleen maar lelijke mensen op Tinder ziet, dat dan wel de mensen zijn die jou aantrekkelijk vinden!

De interessantste ontwikkeling gaat over zogenaamde neurologische netwerken. Die bestaan uit drie lagen:

1. Een input - layer;
2. Een hidden - layer;
3. Een output - layer.



Over de input layer  
Zo'n netwerk noem je een kunstmatig neurologisch netwerk omdat het (een beetje dus, zie eerdere opmerkingen) is gebouwd als onze hersenen. Je hebt synapsen en neuronen en feedback en training. Er zijn heel veel varianten en er zit heel veel techniek achter, maar wij proberen het zo simpel mogelijk uit te leggen. Stel, je wil zo'n netwerk 'trainen' om katten en honden te herkennen. Dan label je miljoenen plaatjes met hond of kat, en dan voer je die in. Het netwerk gaat dan 'kijken' naar de pixels. Zie ik snorharen? Met welke waarschijnlijkheid? Zie ik een vacht? Groene ogen? En zo duizenden factoren en het netwerk verzint er zelf allerlei factoren bij en dan komt er een conclusie (ja, een kat!) en die wordt gevalideerd en zo gaat het netwerk leren. Na enige tijd en heel veel plaatjes is het bijna perfect in het herkennen van katten en honden.

Er zijn wel een paar issues aan te wijzen. Ten eerste, zoals al gezegd, ik hoefde mijn dochter (godzijdank) geen miljoenen plaatjes te laten zien van honden en katten. Dat kon ze veel sneller. Dus werkt het wel als onze hersenen? Als tweede heb je input nodig. Oftewel trainingsdata. Het nadeel van die data is dat ze vaak uit het verleden komt, en dus bevoordeeld kan zijn. **Simpel gezegd, als je een nieuwe toekomst wil maken, moet je niet trainen met data uit het verleden.** Natuurlijk zijn er ook fantastisch veel prachtige toepassingen, waarbij vooroordelen niet of minder gelden. Als derde is er vooral veel data op het internet, en dat is niet altijd een geweldige plek. Twee voorbeelden: [Tay](https://gizmodo.com/here-are-the-microsoft-twitter-bot-s-craziest-racist-ra-1766820160) en [Solid Gold Bomb](https://money.cnn.com/2013/03/05/smallbusiness/keep-calm-and-carry-on/index.html) (Keep Calm and Rape On!)

Over de hidden layer  
Het bovenstaande was al best interessant, vond ik, maar kunstmatige neurologische netwerken worden pas echt interessant wanneer we ons focussen op de verborgen lagen. Ik vind het ook altijd intrigerend, dat op elk plaatje deze lagen ook 'hidden layers' genoemd worden. Geen descriptive layers of definition layers, nee hidden. Een kunstmatig neurologisch netwerk doet er niet geheimzinnig over, dat er iets geheimzinnings gebeurd. En dat is ook zo.

De beste manier om dat uit te leggen is middels het voorbeeld van de Gaydar. Het idee van een Gaydar (Gay Radar) is dat homoseksuele mannen beter in staat zijn om andere homoseksuele mannen te herkennen, dan heteroseksuele mannen dat kunnen. Wanneer je dit wetenschappelijk test door middel van foto's dan blijkt dit niet te bestaan, Dus, dacht Michael Kosinski, een data-wetenschapper, misschien kan ik een AI bouwen die het wel kan. Hij gebruikte 'goedkope' AI, kwam een heel eind en schreef er een [waarschuwende paper](https://qz.com/1078901/a-stanford-scientist-says-he-built-a-gaydar-using-the-lamest-ai-to-prove-a-point/) over, die een hoop onrust veroorzaakte. Het meest verontrustende van een dergelijk systeem is niet alleen **dat** het (misschien) kan, maar vooral dat je vervolgens niet uit kan leggen, **hoe** het systeem het gedaan heeft.

Als je dat wel zou kunnen uitleggen, dan wordt de conclusie van de AI weer onbetrouwbaar. Er is dus een trade-off tussen nauwkeurigheid en transparantie. Dat is in sommige gevallen fantastisch (als een AI op zoek gaat naar huidkanker) en in sommige gevallen zorgwekkend (als een AI bepaalt of je een huis toegewezen krijgt).

Een stukje eerder in deze long read hebben we gesproken over de schaakcomputer AlphaZero van Google/Alphabet. Deze computer heeft een broertje genaamd AlphaGo. In 2015 versloeg AlphaGo één van de beste GO - spelers in de wereld. Dat was knap want Go is een spel dat vele malen complexer is dan schaak en vooral gespeeld wordt met creativiteit en intuïtie (op gevoel). En toch won AlphaGo. Opvallend was dat in één wedstrijd AlphaGo een zet deed die nog nooit vertoond was. Een zet die geen mens zou maken. Iedereen vond het prachtig. Creatief. Intuïtief. Maar het was ook vreemd. Want **waarom** deed AlphaGo die ze? De [uitleg](https://www.wired.com/2016/03/googles-ai-viewed-move-no-human-understand/) van de engineer was beperkt. Snappen wij waarom neurologisch geprogrammeerde systemen tot bepaalde conclusies komen? En als we dat niet snappen, hoe zit het dan als ze geen spelletjes spelen maar besluiten over onze banen, leningen, behandelingen of straf? Isaac Asimov beschreef de drie wetten van robotica. De vraag is of daar geen vierde aan toegevoegd moet worden: Leg uit!. Een robot moet uit kunnen leggen waarom hij tot een bepaalde conclusie komt.

(Lokale) overheden maken steeds vaker gebruik van een Automatic Decision System (ADS). Dit is een systeem, met kunstmatige intelligentie, dat automatisch een besluit kan nemen. Bijvoorbeeld of je wel of geen uitkering krijgt, of een gevangenisstraf of een huurwoning. Er zijn een aantal problemen met dit soort systemen. Ten eerste is de kans groot dat er vooroordelen in het ADS zitten. Immers, de trainingsdata zit ook vol met vooroordelen. Een manier die je vaak hoort is om dit voorkomen met AI-principes (deze hebben verschillende namen zoals Fairness of Fate). Sommige bedrijven, zoals IBM, hebben dat zelfs [geautomatiseerd](https://www.ibm.com/blogs/research/2018/09/ai-fairness-360/) (link naar site). Het idee is dat je AI bouwt die ongewenste vooroordelen uit de beslissing haalt, die transparant is en verantwoordelijkheid neemt.

Hier een [mooi overzicht](https://www.visualcapitalist.com/every-single-cognitive-bias/) van allerlei soorten (menselijke) vooroordelen.  
Hier een [PDF](http://technofilosofie.com/wp-content/uploads/2016/12/FPF-Automated-Decision-Making-Harms-and-Mitigation-Charts.pdf) over issues met ADS en mogelijke oplossingen.

Makkelijk is dat niet want:

* Hoe train je een AI met data als data bevooroordeeld is?
* Welke vooroordelen zijn gewenst en welke niet? Bijvoorbeeld vrouwen zijn veel minder vaak recidivist, moet je die er dan uithalen of juist niet? Sommige vooroordelen bevorderen de eerlijkheid.

En zelfs als je het ADS eerlijk hebt gemaakt (als dat al kan), dan betekent nog niet dat de sociale impact gewenst is. Stel bijvoorbeeld dat de ADS onderdeel is van een systeem dat niet of niet eerlijk functioneert, houdt de efficiency van de ADS dan niet juist het systeem in leven? Daarnaast kijken de AI - principes vaak naar de ADS, maar ze zouden moeten kijken naar een bredere impact. Bijvoorbeeld in Boston werden allerlei van vooroordelen ontdane data ingevoerd in een ADS. Dat ADS probeerde op basis van leeftijd, daglicht, bustijden en schoolspreiding een ideaal schema te maken met flexibele starttijden voor schoolkinderen. Dit lukte nog ook. Het schema zorgde dat het minder druk was in de bussen en dat schooltijden meer aangepast waren op het ritme van de kinderen. Het ADS vergat echter dat ouders met minder inkomen vaak veel minder flexibele roosters hebben en minder geld voor kinderopvang. Daarmee was het ADS, hoe eerlijk ook, erg oneerlijk in zijn impact.

Een manier om naar de hele impact te kijken en de juiste stakeholders te betrekken is om gebruik te maken van het [AIA - FrameWork van AI NOW](https://ainowinstitute.org/aiareport2018.pdf). (link naar PDF).

Er is een hoop discussie over 'de duisternis' rond AI. Vaak valt de term explainable AI, maar die wordt op allerlei manieren gebruikt. Het nadeel van AI die je kunt uitleggen is, zoals we al gezien hebben, dat het ten koste gaat van de nauwkeurigheid. Het nadeel van het bouwen van een AI, die een andere AI kan uitleggen, is dat de vraag opdoemt, en wie legt dan weer de AI uit, die de AI uitlegt. Het laatste scenario begint te lijken op DeepMind, de beroemde computer uit de HitchHikers Guide To the Galaxy, die na lang rekenen als antwoord op de vraag wat is de zin van het leven, concludeerde: 42. En toen mensen zich afvroegen, maar wat betekent dat antwoord, suggereerde DeepMind om een betere computer te bouwen, die de vraag kon herformuleren.

Een ander punt in deze discussie is dat veel mensen pleiten voor een 'human-in-the-loop'. Het idee dat een mens altijd meekijkt. Persoonlijk vind ik dat een beetje mager. Er zit ook een zekere hiërarchie in, namelijk eerst de computer en daarna de mens. Misschien zou daarom een betere term zijn: computer-in-de-loop. Dat klinkt gezonder.

**(3) Zomaar wat gedachten over AI**

Hieronder een aantal losse gedachten over de impact die AI zou kunnen hebben, de komende jaren in onze maatschappij. Laten we beginnen met wat (min of meer) positieve dingen.

A. Dingen die 'denken'  
Volgens Kevin Kelly in zijn boek [The Inevitable](http://kk.org/books/the-inevitable/)(link naar boek) kun je vergelijkingen trekken tussen de opkomst van**electriteit en kunstmatige intelligentie**. Je neemt een apparaat, en voegt er iets aan toe. Bijvoorbeeld, je had lang geleden een handboor, electriciteit erbij, hoppa, een electrische boormachine. Je had een tandenborstel, electriciteit erbij en hoppa een electrische tandenborstel. Nu voeg je weer wat toe: kunstmatige intelligentie. Dat leidt opnieuw tot een nieuw product. Bijvoorbeeld, de electrische tandenborstel, AI erbij, en hoppa, je hebt een mini-tandarts die elke dag ook even je gebit controleert en suggesties geeft. Eerst wassen met de hand, toen een een electrische wasmachine en straks een machine waarbij je per was betaalt afhankelijk van de beschikbare stroom op het slimme grid. Zo kan elk denkbaar product voorzien worden van kunstmatige intelligentie. Het gaat dan vooral om kunstmatige intelligentie die uit 'een cloud' komt en heel specifiek 'iets' kan. Er zal, zo voorspelt Kevin Kelly, juist geadverteerd worden met kunstmatige intelligentie die heel goed is in één ding. Je wil niet dat je zelfrijdende auto tijdens het rijden ook nog probeert een boekhoudkundig probleem op te lossen. Maar Kevin Kelly is dan ook een onverbeterlijke optimist.

B. Dingen die tot leven komenIn het meest extreme geval, las ik een verhaal in het [NRC](https://www.nrc.nl/nieuws/2018/11/23/algoritmen-kunnen-toveren-a2756353?utm_source=SIM&utm_medium=email&utm_campaign=Gespreksstof&utm_content=&utm_term=20181212) (link naar artikel) over een wereld waarin alle dingen weer tot leven komen. Zeg maar terug naar het animisme, het geloof dat alle dingen op aarde krachten bezitten. Denk aan bomen die praten tegen het gras, het gras dat dat doorverteld aan de bloemen, die jij plukt, maar niet voordat ze 'au' hebben gezegd. Wazig? Lezen dat stuk!

C. Ondersteunen bij besluitvorming  
Kunstmatige intelligentie kan mensen ondersteunen in de besluitvorming. Je ziet dit heel vaak in de medische zorg. Kunstmatige intelligentie is veel beter en sneller in staat om röntgenfoto's te bekijken, om databases met symptomen en behandelingen door te ploegen, om verbanden te leggen. Er zijn bijvoorbeeld veelbelovende doorbraken op het gebied van borstkanker waarbij kunstmatige intelligentie andere patronen ziet en zo de dokter kan helpen. De computer van IBM, genaamd Watson, wordt daar nu veelvuldig voor ingezet. Nu is het natuurlijk al zo dat een simpele Google Search (die onder de motorkap verre van simpel is) je kan helpen om tot een beter besluit te komen, mits je weet wat je doet. Dat zullen we bij steeds meer processen zien en op een steeds geavanceerde manier. Een semi-zelfrijdende auto combineert dingen die denken en tot leven komen met besluitvorming. Uiteraard draagt ondersteunen bij besluitvorming ook heel erg bij aan meer efficiency.

D. Efficiency  
Als we het dan toch over efficiency hebben, dan kunnen we ook wat verder kijken dan alleen de toepassing in het bedrijfsleven. Denk ook eens aan een AI die jou leert kennen, die je patronen weet, die jou begrijpt en met die kennis doet de AI de boodschappen. De AI winkelt, zodat jij alle tijd hebt om je te storten op slow cooking. En zo zijn er veel voordelen te bedenken.

E. Je Persoonlijke CoachEr komt steeds meer informatie, dus zijn er overal algoritmes die voor jou een voorselectie maken, zodat je niet door alle informatie heen moet. Daar is veel over te zeggen (zie verderop), maar in essentie is het handig. Een systeem dat alvast voorsorteert kan je tijd besparen en zorgen dat de tijd die je over hebt nuttig besteed wordt. Hetzelfde zien we op andere vlakken. Bijvoorbeeld kunstmatige intelligentie die bij houdt hoe je beweegt, slaapt, adem haalt, je hart slaat, etc... en met die informatie goedbedoelde adviezen geeft. Of kunstmatige intelligentie die met sensoriek je ochtend-ontlasting checkt en op basis daarvan adviezen geeft van ga naar de dokter tot iets meer drinken.

En zo zijn er nog véél en véél meer toepassingen te bedenken van kunstmatige intelligentie die helemaal niet zo vergezocht zijn. Het hoeft niet meteen te gaan over killerbots of AI die slimmer is dan mensen.

En dan zijn er ook nog leuke dingen, zoals de opstart [No More Woof](http://www.nomorewoof.com/) (link naar site) die beweerde 14 emoties in je hond te kunnen herkennen door met AI hersengolven te analyseren. Je hoeft dan niet meer te denken: wat ziet mijn tackle er treurig uit. Nee, je weet het zeker. De opstart kreeg ontzettend veel aandacht, haalde heel veel geld op middels crowdfunding. En daarna werd het stil. Een mooi voorbeeld van het geloof in AI, en het gegeven dat intelligentie weliswaar steeds vaker kunstmatig is, maar niet alom aanwezig!

Er zitten natuurlijk ook negatieve kanten aan de snelle ontwikkeling van geavanceerde kunstmatige intelligentie. Die zetten we hier op een rij.

A. Dingen die zich overal mee bemoeien  
Je hebt ook natuurlijk dingen die denken, en zich vervolgens overal mee bemoeien. Mijn favoriete voorbeeld is de [HapiFork](https://www.bright.nl/tests/artikel/3975277/uitpakparty-hapifork) (artikel op Bright). Dat is een vork die bijhoudt hoe snel je eet (want snel eten is slecht voor je!) en vervolgens gaat die vork trillen als je te snel eet, zodat het moeilijk wordt om je vlees aan te prikken. Zo krijg je wellicht ook stoelen die bijhouden of je rechtop zit, auto's die in de gaten houden of je wel goed rijdt, boeken die weten of je snel genoeg en genoeg leest, koffie-automaten die je caffeïne inname controleren en ga zo maar door. Voor mensen die al een hekel hebben aan het pingende lichtje in de auto, breken slechte tijden aan.

In het algemeen zou je kunnen zeggen dat dingen die je autonomie vergroten (die je betere beslissingen laten nemen) een vooruitgang zijn en dingen die je autonomie verkleinen (die je dwingen bepaalde handelingen te verrichten), zijn dat juist niet.

B. Processen die worden geoptimaliseerd  
Een extreem geval van de aantasting van analytics zijn bedrijven die werken onder de eufemistische naam People Analytics of Workforce Analytics. Het idee is dat er data verzameld wordt over hoe mensen werken. Dat kan zijn hoe mensen typen of hoe mensen werken in een magazijn (bijvoorbeeld door ze armbanden te laten dragen) en die data vervolgens te analyseren met AI. Daarna kun je die data weer gebruiken om processen te optimaliseren. Er is best iets voor te zeggen als je baas bijvoorbeeld kijkt naar de geanonimiseerde of metadata van alle mailverkeer en daar een analyse van maakt. Het wordt een stuk enger wanneer je baas kijkt naar jouw mailgedrag. Of kijkt of jij wel fanatiek genoeg typt. Of genoeg stappen zet in het magazijn. In dat laatste geval hoef je niet bang zijn dat je vervangen wordt door een robot, maar wel dat je behandeld wordt als een robot. Het verschil met vroeger is dat de techniek er rijp voor is, en dat betekent dat de discussie over wat je ermee doet nu echt belangrijk wordt.

En dit doembeeld kan nog veel verder gaan. Wat dacht je namelijk van zogenaamde algoritmes van vlees (meat algorithms). Lees er [hier](http://technofilosofie.com/ai-meer/) meer over.

C. GezichtsherkenningEen ander voorbeeld waar mensen zich op dit moment serieus zorgen om maken is gezichts,- en emotieherkenning. De apparatuur om gezichten te herkennen, met behulp van AI, wordt steeds beter en daarmee wordt het ook steeds beter mogelijk om emoties te herkennen. Zo kun je zien of je wel oplet in de klas. Of, op een callcenter, zien de callcenter-medewerkers op een scherm de emoties van de beller. Er worden zelfs weer eigenschappen gekoppeld aan gezichtsuitdrukkingen en aan emoties. Lees ook dit [artikel](https://theintercept.com/2018/12/06/artificial-intellgience-experts-issue-urgent-warning-against-facial-scanning-with-a-dangerous-history/) op Intercept dat verwijst naar de lessen uit het verleden én pleit voor regels.

Kunstmatige intelligentie kan automatisch gezichtsuitdrukkingen analyseren, op enorme schaal en er kunnen voorspellingen aan gekoppeld worden. Er wordt zelfs steeds vaker gebruik gemaakt van [physiognomie](https://en.wikipedia.org/wiki/Physiognomy) (link naar de Wiki). Physiognomy is de pseudo - wetenschap die stelt dat karaktereigenschappen kunnen worden gekoppeld aan uiterlijke kenmerken. Ondanks dat hier geen bewijs voor is. Ook wordt steeds vaker gesproken of gebruik gemaakt van emotieherkenning, ondanks dat dit geen betrouwbare technologie is. Ook gezichtsherkenning is niet betrouwbaar. Zwarte mensen zijn moeilijker te lezen dan witte mensen, bijvoorbeeld. Daarnaast zijn voice-assistenten als Siri en Alexa druk bezig met emotieherkenning, net als sommige callcenters. Bijkomend probleem is dat deze technologie vaak niet goed werkt, dat de technologie wordt aangeboden door commerciële partijen, die weigeren om de code transparant te maken (want bedrijfsgeheim) en dat ze gebruikt worden door (lokale) overheden.

Vergeet niet, het is nooit wetenschappelijk bewezen dat de leugendetector werkt. Het is wel wetenschappelijk bewezen dat de leugendetector niet werkt. De leugendetector wordt nog steeds op grote schaal gebruikt!

AI Now stelt dat overheden gezichts,- en emotieherkenning ver van zich moeten werpen én dat er strenge regelgeving moet komen. Een bijkomend idee is dat een overheid van nature vaak stelt dat mensen niet te vertrouwen zijn. Immers, onbetrouwbare medemensen vragen om een sterke overheid. Met gezichtsherkenning, die goed werkt, is de kans groot dat de balans tussen overheid en burgers doorslaat in het voordeel van de overheid.

**(4) Moeten we AI reguleren en hoe doen we dat dan?**

Bij de voorbeelden hierboven, doemt steeds dezelfde discussie op. Dankzij AI is er steeds meer mogelijk, maar omdat er steeds meer mogelijk is, is er ook steeds meer behoefte aan een goede discussie. Aan wetgeving. Aan regelgeving. Aan nieuwe normen en waarden. Aan ethiek. **En dat blijft achter**.

Allereerst is het van belang om te weten wat ethiek nu eigenlijk is. Een goede manier is om gebruik te maken van deze [website](https://www.scu.edu/ethics/ethics-resources/ethical-decision-making/) (link naar website) van het Markulla Center waarin heel veel wordt uitgelegd rond ethiek. Of lees de samenvatting op deze [PDF](https://www.scu.edu/media/ethics-center/resources/making.pdf). In het [AI NOW Report van 2018](https://ainowinstitute.org/AI_Now_2018_Report.pdf) (link naar PDF) wordt ingegaan op de ethische aspecten van AI en worden allerlei aanbevelingen gedaan. Samengevat zijn de volgende zaken van belang:

Het is van belang om te beseffen, zoals hierboven al een paar keer aangestipt, dat kunstmatige intelligentie allerlei aspecten van ons leven gaat veranderen. Je kunt je misschien prima voorstellen (of misschien hoop je er zelfs op) dat straks een 'bot' de vergadering gaat leiden. Kun je je ook voorstellen dat een bot straks het gesprek gaat leiden aan de eettafel? Of dat de 'bot' het gesprek met je kinderen zelfs overneemt? Nog wat losse opmerkingen die het belang van ethiek onderstrepen:

* AI kan fantastisch zijn om je te helpen, bijvoorbeeld met je gezondheid of de juiste keuze voor wat je wil zien op je streaming kanaal. Maar als je beter geholpen wilt worden, moet je meer van je zelf blootgeven en daar liggen privacy - issues op de loer;
* Kunstmatige intelligentie maakt bullshitjobs - (zie bij [banen](http://technofilosofie.com/de-toekomst-van-onze-banen/)) misschien overbodig;
* Kunstmatige intelligentie wordt voornamelijk gebouwd door de grote technologiebedrijven. Zij beschikken over de kennis, de trainingsdata, de infrastructuur en het geld. Leidt AI dan ook tot een verdere concentratie van macht?
* 2018 was een dramatisch jaar voor de ethiek van kunstmatige intelligentie. Denk aan de schandalen rondom Facebook. Denk aan het defensie - drone project van Google in het project Maven. Denk aan de keren dat Watson, de kunstmatige intelligentie dokter van IBM het mis had. Denk aan al die lokale overheden zoals de NYPD die projecten starten met gezichtsherkenning en discriminatie. Denk aan China met zijn Sociale Creditscore. En nog veel meer.

Wel zien we dat grote AI - bedrijven steeds vaker zelf nadenken over eigen ethische normen en waarden. Maar zitten we daar eigenlijk wel op te wachten? Immers:

* Technologische bedrijven zien technologische vooruitgang als onvermijdelijk. Ze zullen dus niet gauw tot de conclusie komen dat het beter is om geen AI in te zetten;
* Ethische principes conflicteren vaak met andere principes die gelden voor beursgenoteerde bedrijven (zoals winstmaximalisatie);
* Technologische bedrijven zijn technisch, en hebben dus vaak een beperkte blik. Problemen definieer je zo, dat je ze kunt oplossen met techniek;
* En we hebben al gezien dat ondanks ethische uitgangspunten er toch anders gehandeld wordt. Google houdt zich bijvoorbeeld aan de rechten van de mensen en werkte (of werkt... dat is onduidelijk) aan een gecensureerde zoekmachine voor de Chinese markt (Dragonfly).

Hoe dan ook, de zorgen over KI zijn er wel degelijk en steeds meer mensen pleiten voor een ethische gedragscode. Daarbij laten ze zich inspireren door voorbeelden uit de nucleaire geschiedenis (nucleaire kennis is goed afgeschermd) of gifgas - afspraken (gifgas is makkelijk te maken, maar wordt door de internationale gemeenschap streng verworpen). Zij hebben op de Asilomar - conferentie [23 principes](https://futureoflife.org/ai-principles/?cn-reloaded=1) geformuleerd (link naar principes). De kern van de principes is, zoals altijd, dat de mens, en onze menselijke waarden, centraal staan in het denken.

Lijkt mij een goed idee.

**(5) Krijgen machines straks verbeeldingskracht?**  
Een interessante, vind ik, nieuwe ontwikkeling zijn de Generative Adverserial Networks. Simpel gezegd zijn het computers met verbeeldingskracht. Het zijn twee kunstmatige neurologische netwerken die met elkaar praten. De één verzint iets (de generator) en de ander controleert hetgeen verzonnen is (de discriminator). Op die manier kun je heel snel, heel veel variaties verzinnen op thema's.

Gave voorbeelden zijn:

1. [Thispersondoesnotexist.com](https://thispersondoesnotexist.com/) - een website, die elke keer als je op F5 drukt nieuwe mensen verzint;
2. Een [animatie](https://www.fastcompany.com/90352523/oh-no-samsung-can-create-a-video-of-you-from-a-single-still-photo) vanuit één foto;
3. DeepNude en aanverwante [programma's](https://www.ad.nl/tech/opnieuw-een-naaktfoto-app-op-de-markt-nadat-deepnude-offline-werd-gehaald~af63be26/) - die foto's van vrouwen omzetten in (verzonnen) naaktfoto's van vrouwen;
4. [Bot Dylan](https://consequenceofsound.net/2017/05/meet-bot-dylan-the-ai-computer-that-can-write-its-own-folk-songs/), die zelf liedjes kan schrijven;
5. [Bot Ross](https://towardsdatascience.com/bot-ross-teaching-a-computer-to-paint-f5a9c9ef908), die zelf schilderijen maakt van een paar ruwe componenten;
6. En [Talk To Transformer](https://talktotransformer.com/)die zinnen afmaakt;

En vind veel meer voorbeelden in [deze PDF](https://www.jarnoduursma.nl/machines-met-verbeeldingskracht-een-kunstmatige-realiteit/) van Jarno Duursma.

Er zijn een heleboel consequenties en gevolgen te bedenken rondom deze ontwikkeling. Ik beperk me tot 3 mogelijke gevolgen:

1. Fake News;
2. Het meester - gezel - model;
3. De opkomst van het échte.

GAN's bieden in eerste instantie veel mogelijkheden als jij iemand bent die Fake News produceert. We hadden al last van Cheap Fakes, waarin geknipt wordt in video of video's vertraagd worden afgespeeld, maar de opkomst van de Deep Fakes is onstuitbaar. Maar als het heel gemakkelijk wordt om Fake News te maken, zou dat ook het einde kunnen betekenen van Fake News. Misschien vinden we straks dat alle nieuws Fakes is, behalve True News en wordt dat de nieuwe term.

Als tweede kennen we heel veel beroepen waarin je van onderaf moet opklimmen. Stel, je bent een architect, dan teken je eerst de schuurtjes, of de initial drafts en dan leg je die voor aan de senior architect, die keuzes maakt. Hetzelfde geldt voor advocaten, die beginnen met papierwerk, of designers, die eerst zoveel mogelijk schetsjes maken, of animators die beginnen aan achtergronden. Maar wat als dat nu allemaal gedaan wordt door GAN's? Kun je het vak nog leren, zonder al die uren te stoppen in het grondwerk? Als er geen gezellen zijn, ontstaan er dan nog meesters?

En tenslotte, als alles op het web eenvoudig te faken is, komt er dan een opleving van het echte. Ik denk het wel. Immers, online is alles te faken. Ook als je supergoed kunt dansen, fantastisch kunt zingen, prachtig de bal kunt hoog houden, er lekker kunt uitzien! Daarom zullen we steeds meer dingen in het écht willen zien. Kopen dus dat oude theatertje!

**(4) Moeten we ons zorgen maken om onze banen?**Een discussie die vaak gevoerd wordt, is of we onze zorgen moeten maken om onze banen als gevolg van AI. We hebben daar een aparte pagina aan gewijd waarin we deze discussie bijhouden. Die vind je [hier](http://technofilosofie.com/de-toekomst-van-onze-banen/).

**BIJLAGE A. EXTRA ACHTERGROND INFORMATIE**

Op deze pagina bieden we achtergrond - informatie aan rondom het onderwerp kunstmatige intelligentie, maar laten we eerlijk zijn. Even Googlen/DuckDuckGo-en werkt ook.

Daarnaast bieden we lesmateriaal aan onder deze [link](https://maken.wikiwijs.nl/136737/Technofilosofie___Aandacht#!page-5010291) en de masterclasses onder deze [link](https://maken.wikiwijs.nl/136737/Technofilosofie___Aandacht#!page-5633958).

Ten eerste adviseren we je om het boek van Max Tegmark, [Life 3.0](https://www.bol.com/nl/f/life-3-0/9200000078814497/). Te lezen. Max noemt het zelf de Survivalgids voor de toekomst, maar dat is waarschijnlijk onder druk van de marketing afdeling. Het is voornamelijk een uitgebreid, evenwichtig en positief boek over kunstmatige intelligentie van iemand die midden in die wereld staat. Als je meer wilt weten over kunstmatige intelligentie, en voorbij de hype mee wil praten of denken, is dit boek een aanrader. Beetje saai, wel heel goed. Michio Kaku gaat juist in op de toekomst van de geest in zijn boek [The Future of Mind](https://www.bookdepository.com/The-Future-of-the-Mind-Department-of-Physics-Michio-Kaku/9780385530828?redirected=true&utm_medium=Google&utm_campaign=Base1&utm_source=NL&utm_content=The-Future-of-the-Mind&selectCurrency=EUR&w=AF74AU9SC1ZT57A80RT3&pdg=aud-346056386285:pla-311561233428:kwd-311561233428:cmp-869901575:adg-43267140629:crv-203874775887:pid-9780385530828:dev-c&gclid=CjwKCAjwhqXbBRAREiwAucoo-wH5gvIFtraF7DRCAwtFpv9f0FHNGKWqmtpsJ24RCXtftdxYUehtmRoCZbMQAvD_BwE). Soms wel heel technisch maar vol interessante beelden en een toekomst met heel andere brain-machine interfaces. Ik heb hem ook wel eens zien presenteren. Mijn conclusie: hij praat grappiger, dan hij schrijft!

Wat hebben we nog meer:

* Een interessante [whitepaper](https://www.pwc.co.uk/audit-assurance/assets/explainable-ai.pdf) van PwC over explainable AI. Niet makkelijk, wel goed.
* Uitleg over de AI in de [nationale AI Cursus](https://app.ai-cursus.nl/home) van Jim Stolze. Leuk, informatief, toegankelijk!
* Een [website over de toekomst van AI](http://www.detoekomstvan.ai/)met een interessant rapport (97 pagina's) over AI. Aanbevelingswaardig!
* Daarnaast zijn interessante links <https://www.robot-onderwijs.nl/> met vooral lesmateriaal voor basisscholen en alle nieuwe ontwikkelingen kun je volgen via het [Future of Life Institute](https://futureoflife.org/).
* Ook wijzen we je op dit [blog over internet - ethiek](http://straks:%20https/www.scu.edu/ethics/internet-ethics-blog/) met veel AI;
* Of dit artikel van Luciano Floridi op Aeon over de [toekomst](https://aeon.co/essays/true-ai-is-both-logically-possible-and-utterly-implausible) van AI
* Het is een beetje oefenen en het is een beetje ervaren. Deze [film](https://www.innovatiefinwerk.nl/mens-en-robot-2045) van innovatief-in-werk met extra, verdiepende vragen.

BIJLAGE B. GENIETEN VAN AI

Op deze pagina staan grappige, trieste, interessante en leuke dingen rondom AI.

Check deze website vol Vreemde Kunstmatige Intelligentie: [aiweirdness.com](https://aiweirdness.com/)

**EEN VAL VOOR ZELFRIJDENDE AUTO'S**  
Eén van de onderwerpen die altijd snel naar voren komt als het gaat over kunstmatige intelligentie zijn de zelfrijdende auto's. Maar hoe slim zijn die dingen eigenlijk. Kunstenaar James Bridle bouwde een val.

<https://www.youtube.com/watch?v=thuN2HD6m2s>

**TAY THE RACIST CHATBOT**

In maart 2016 lanceerde Microsoft een chatbot voor twitter. Op haar profiel stond Tay Ai. De kunstmatige intelligentie imiteerde een tienermeisje waar je mee kon twitteren. Het idee was, hoe meer je praat, hoe slimmer Tay zou worden. Helaas, binnen 24 uur, ging Tay 'Full Goebels'. Andere twittergebruikers waren domme hoeren, negers moesten in concentratiekampen, de Holocaust werd ontkend, een Mexicaanse volkerenmoord aangemoedigd. Triest of grappig, of beide. Maar er zijn wel een paar belangrijke lessen te leren.

* Is dit een probleem van de ontwerpers of van de Twitter trollen? Had dit voorkomen kunnen worden? Hoe dan? Zijn er legale consequenties? Wat betekent het als Tay 100.000 Tweets per dag afvuurt? Hebben we inzicht in hoe Tay tot haar Tweets kwam?

Lees daarom dit [artikel](https://futureoflife.org/2016/03/27/tay-the-racist-chatbot-who-is-responsible-when-a-machine-learns-to-be-evil/), dat ook alle andere artikelen rondom dit onderwerp opsomt.

**KEEP CALM AND RAPE ON**

Ander voorbeeld. Gold Bomb, een Australisch bedrijf dat shirts verkocht met opdruk. Je kon uiteraard zelf opdrukken maken, maar er waren ook standaard opdrukken, en die werden bepaald door een algoritme dat uitzocht wat populair was op het internet. Deze standaard-opdrukken verschenen in de catalogus van Amazon.com en waren, op zijn zachtst gezegd, bizar, maar wel een uiting van een probleem waar veel mensen zich zorgen over maken. De AI vormt zich op basis van trainingsdata, die is vaak te vinden op het internet, en wat vinden we eigenlijk van de 'toon' van het internet?

**THE NIGHTMARE VIDEOS OF CHILDREN'S YOUTUBE**

Fantastische Ted Talk van James Bridle, de auteur van New Dark Age, over YouTube en het internet/KI in het algemeen.

<https://www.ted.com/talks/james_bridle_the_nightmare_videos_of_childrens_youtube_and_what_s_wrong_with_the_internet_today>

#### JOHN OLIVER OVER ROBOTISERING

Komische video van held John Oliver over Robotisering (19.43 minuten)

<https://www.youtube.com/watch?v=_h1ooyyFkF0&amp;app=desktop>