Griezelig slim

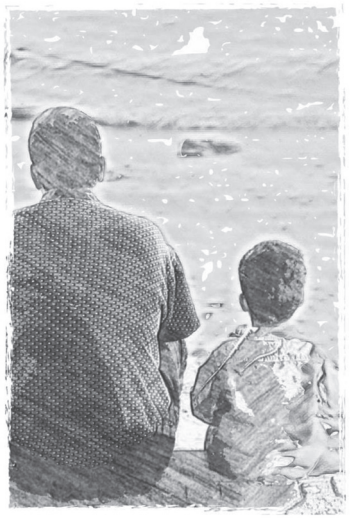
De impact van kunstmatige intelligentie op ons leven

Mo Gawdat

Amsterdam 2021
Uitgeverij Brandt

The gravity of the battle means nothing

to those at peace



Voor Ali
It's now or never
It's me and you

Inhoud

Inleiding: De nieuwe superheld 9

Deel een: Het griezelige deel

1 Een korte geschiedenis van intelligentie 29
2 Een korte geschiedenis van onze toekomst 53
3 De drie onvermijdelijkheden 71
4 Een milde dystopie 104
5 De touwtjes in handen 139

Deel twee: Onze route naar Utopië

6 En zij leerden 175
7 Onze toekomst opvoeden 205
8 De toekomst van de ethiek 233
9 Vandaag heb ik de wereld gered 261

De Universele Verklaring van de Rechten van de Wereld 321
Nawoord 325
Noten 339

Inleiding

De nieuwe superheld

Dit boek is je alarmbel. Het is geschreven voor iedereen die niet goed is geïnformeerd over de pandemie die eraan zit te komen: die van de kunstmatige intelligentie (ook wel *artificial intelligence* of AI genoemd). Het zal worden bekritiseerd door deskundigen en dat is precies de reden waarom ik het schrijf. Want om expert in kunstmatige intelligentie te worden moet je er een deskundige, nauwe blik op nahouden. Die blik gaat volledig voorbij aan de existentiële aspecten van kunstmatige intelligentie, die los staan van de technologie: kwesties als moraal, ethiek, emoties, compassie en een hele reeks ideeën waar filosofen, spiritueel zoekenden, idealisten, milieuactivisten en – meer in het algemeen – gewone mensen (dat wil zeggen: wij allemaal) zich mee bezighouden. Bovendien is het belangrijkste uitgangspunt van dit boek dat het je wil laten zien dat *niet* de experts het gevaar kunnen afwenden van de superintelligentie die de mensheid bedreigt. Nee, jij en ik hebben die macht. En belangrijker nog: jij en ik dragen die verantwoordelijkheid.

Rond de tijd dat dit boek verschijnt sluiten we een periode van twee jaar af waarin we met de coronapandemie hebben geleefd. We zijn optimistisch gestemd omdat de vaccins hun vruchten beginnen af te werpen en de kans

bestaat dat ons leven weer normaal wordt. Maar wat 'normaal' is, is voortdurend aan verandering onderhevig. Ik denk dat de manier waarop de wereldgemeenschap en onze politieke leiders de uitbraak van corona hebben aangepakt niet zo heel veel verschilt van de manier waarop ze zullen omgaan met de naderende uitbraak van de kunstmatige-intelligentiepandemie. Ik hoop maar dat ze leren van de fouten die ze in het geval van corona hebben gemaakt en deze volgende verandering zo zullen aanpakken dat die minder diep ingrijpt, beter valt te voorspellen en minder nadelen heeft voor de maatschappij en de economie.

Laat je alsjeblieft niet misleiden door de eenvoud die ik in dit boek nastreef. De feiten die mijn beweringen ondersteunen vallen niet te ontkennen. Die ontleen ik aan mijn lange loopbaan van meer dan dertig jaar in de techsector. Vóór mijn huidige start-up (die op een dusdanige manier gebruikmaakt van geavanceerde computersystemen, robotica, kunstmatige intelligentie en technologieën voor machinaal leren dat het denkbaar is dat we er de wereld mee kunnen redden) waren de twaalf jaar die ik bij Google werkte een van de hoogtepunten in mijn carrière. Bij Google had ik het voorrecht verantwoordelijk te zijn voor de bedrijfsvoering en mocht ik technologieën lanceren in bijna de helft van de kantoren van het bedrijf wereldwijd, een gebied waar meer dan honderd verschillende talen worden gesproken. Ik sloot mijn tijd bij Google af als Chief Business Officer van Google [X], de beruchte innovatiepoot waar enkele projecten voor de ontwikkeling van kunstmatige intelligentie werden uitgebroed, zoals zelfrijdende auto's, Google Brain en bijna alle innovaties van Google op roboticagebied.

Mijn inzicht in waar het om draait bij de ontwikkeling

van kunstmatige intelligentie zoals we die vandaag de dag kennen, heb ik deels opgedaan in mijn tijd bij Google [X] en is daarmee uniek. Ik combineer mijn directe ervaring met het onderwerp met mijn werk op het terrein van geluksonderzoek (opgetekend in mijn internationale bestseller *De logica van geluk*, de succesvolle podcast *Slo Mo* en de door mij opgerichte non-profit-organisatie OneBillionHappy.org) om je een uniek perspectief te bieden op de uitdagingen waar we tijdens de dageraad van superintelligentie mee worden geconfronteerd. Ik hoop dat we met behulp van kunstmatige intelligentie een utopie kunnen creëren die de mensheid dient, in plaats van een dystopie die haar ondermijnt. In dit boek bepleit ik dat dat een verantwoordelijkheid is die iedereen moet nemen om een betere toekomst te creëren voor ons allemaal.

Maak je geen zorgen. Dit is geen door angst ingegeven sciencefictionverhaal, eerder een verhaal over een van de grootste kansen voor de mensheid. Dit is een kans om een einde te maken aan onze excessieve afhankelijkheid van consumentisme en technologische vooruitgang, die weliswaar de kwaliteit van ons leven opstuwen, maar dat doen ten koste van elk ander levend wezen op aarde. Alleen als wij het heft in eigen hand nemen en voor verandering zorgen, is dit een hoopvol verhaal.

Ergens ver weg, in een uithoek

Stel je om te beginnen eens een zwakke, bejaarde versie van mij voor die in de wildernis bij een kampvuur zit in het jaar 2055, exact negenennegentig jaar nadat het verhaal over kunstmatige intelligentie is begonnen op Dartmouth Colloge, in New Hampshire. Ik vertel je wat ik sinds de opkomst van kunstmatige intelligentie heb meegemaakt, een verhaal dat ertoe heeft geleid dat we hier in deze uithoek

zitten. Maar ik vertel je pas aan het einde van het boek of we hier zelfvoorzienend leven om uit de klauwen van de machines te blijven of omdat kunstmatige intelligentie ons heeft ontslagen van onze dagelijkse verplichtingen en ons de tijd, de veiligheid en de vrijheid heeft gegeven om van de natuur te genieten, om te doen waar mensen het beste in zijn: verbinden en nadenken.

Ik vertel je dat nu nog niet omdat ik op dit moment gewoon nog niet weet hoe ons verhaal over de machines zal aflopen. Dat, mijn vriend, hangt mede van jou af. Ja, van jou als individu. En niet van je overheid, je baas of de denkers die je volgt. De toekomst hangt daadwerkelijk van jou af. Die wordt bepaald door wat jij de komende tien jaar besluit te doen, te beginnen vanaf vandaag.

Hier volgt een voorspelling. Ik heb in de jaren waarin ik in de voorhoede van de technologie opereerde van dichtbij meegemaakt dat we machines bouwden die slimmer zijn dan wij. Ik heb hoogstpersoonlijk bijgedragen aan de opkomst van de kunstmatige intelligentie. Ik geloofde in de belofte dat technologie ons leven voortdurend zou verbeteren. Totdat ze dat niet langer deed. Toen me de ogen werden geopend, besefte ik dat de technologie, met elke verbetering die ze ons bracht, ook iets van ons afpakte.

Technologie vormt tegenwoordig een ongekend gevaar voor de aarde en al haar bewoners. Dit boek is niet voor de techneuten die de code ervan schrijven, de politici die beweren dat ze haar kunnen beteugelen en de experts die de hype eromheen blijven aanjagen. Die weten allemaal al wat ik je ga vertellen. Dit is een boek voor jou, voor je beste vriend of vriendin of voor je buur. Want, echt waar, wij zijn de enigen die onze toekomst gestalte kunnen geven, maar alleen als we samen het roer in handen nemen en beloven de juiste maatregelen te zullen treffen. Dit boek is een beweging, het begin van een opstand, en die moet snel beginnen omdat we nog maar weinig tijd hebben, hoe graag ik je ook anders zou willen doen geloven. De hoofdstukken van het verhaal dat ik je ga vertellen schrijven we al zeventig jaar. Het wordt tijd dat we allemaal – ook jij – er een einde aan breien.

De nieuwe superheld

Het verhaal over onze toekomst dat jij en ik momenteel schrijven gaat als volgt.

Stel dat een ruimtewezen, compleet met superkrachten, als kind op aarde terecht zou komen. Zonder te zijn grootgebracht met onze aardse waarden kan deze bezoeker zijn krachten gebruiken om onze wereld beter en veiliger te maken, maar hij kan ook een onstuitbare superschurk worden die over de macht beschikt de aarde te vernietigen. Terwijl hij als kind opgroeit, weet hij nog niet welke van beide uitersten hij zal kiezen.

Ik denk dat je het er wel mee eens zult zijn dat het belangrijkste moment voor de toekomst van onze planeet dat is waarop het kind op aarde terechtkomt. Daar hangt vanaf welke ouders het kind zullen vinden, zullen adopteren en de waarden zullen bijbrengen die zijn toekomst bepalen.

In het beroemde gelijknamige superheldenverhaal wordt Superman als kind geadopteerd door Jonathan en Martha Kent. In de meeste verhalen over de eerste jaren van Superman worden ze voorgesteld als liefhebbende ouders die Clark een sterk moreel besef bijbrengen. Ze moedigen hem aan zijn superkrachten te gebruiken om de mensheid te dienen en scheppen daarmee de Superman die wij kennen, degene die ons beschermt en helpt.

Maar het verhaal vertelt nooit hoe Superman zou zijn

opgegroeid als hij agressieve, hebzuchtige en egoïstische ouders zou hebben gehad. Zo'n versie zou waarschijnlijk een superschurk hebben opgeleverd die de mensheid voor eigen gewin te gronde zou hebben gericht.

Het verschil tussen de superschurk en de superman zit hem niet in de superkrachten waarover ze beschikken, maar in de normen en waarden die ze van hun ouders meekrijgen.

Laat me je vertellen dat zo'n met superkrachten uitgerust wezen daadwerkelijk op aarde is geland. Het verkeert nog steeds in het kinderstadium, en hoewel het niet-biologisch van aard is, beschikt het over ongelofelijke eigenschappen. Ik doel uiteraard op kunstmatige intelligentie. In feite is er niets kunstmatigs aan kunstmatige intelligentie; ze is een echte vorm van intelligentie, zij het een die verschilt van de onze.

Kunstmatige intelligentie is nu al slimmer dan ieder mens op aarde als het gaat om allerlei specifieke, opzichzelfstaande taken. Niet lang nadat de computer zijn intrede in ons leven had gedaan, werd een computer wereldkampioen schaken. De kampioen Jeopardy!, een Amerikaanse tv-spelletje, is Watson, een supercomputer van IBM. De wereldkampioen Go heet AlphaGo en is gemaakt door Google. (Go is een abstract bordspel dat ruim 2500 jaar geleden in China werd uitgevonden en vanwege het oneindige aantal mogelijke bordopstellingen als een van de ingewikkeldste strategische spellen wordt beschouwd.) Machines met fabelachtige patroonherkenningseigenschappen vormen het hart van onze veiligheidssystemen, simpelweg omdat ze meer zien dan wij, en verreweg de veiligste chauffeur ter wereld is een zelfrijdende auto die niet alleen verder kan kijken, maar zijn aandacht uitsluitend op de weg houdt. Door gebruik te maken van technologieën met meerdere

sensoren waarmee hij met auto's om hem heen kan communiceren, kan de auto zelfs 'om de hoek kijken'. Als je machines maar genoeg 'traint', in welke vaardigheid ook, leren ze die beter uit te voeren dan wie ook.

Op weg naar het onbekende

Er wordt voorspeld dat machinale intelligentie in 2029, en *dat* is al zo'n beetje om de hoek, de grenzen van specifieke intelligentie zal overschrijden en zal veranderen in algehele intelligentie. Tegen die tijd zullen er dus machines zijn die slimmer zijn dan mensen, punt. Niet alleen worden ze slimmer, ze zullen ook meer weten dan mensen (omdat het geheugen waaruit ze putten het complete internet is) en beter met elkaar communiceren, waardoor hun kennis zich nog verder uitbreidt. Sta daar eens bij stil: als jij een ongeluk met je auto krijgt, dan leer je daarvan, maar als een zelfrijdende auto een fout maakt, dan leren *alle* zelfrijdende auto's daarvan, stuk voor stuk, inclusief de exemplaren die nog niet zijn 'geboren'.

In 2049, nog tijdens ons leven en zeker in dat van de volgende generatie, is kunstmatige intelligentie naar men denkt een miljard keer zo slim (in alles) dan de slimste mens. Om dat in perspectief te plaatsen: jouw intelligentie is vergeleken met die van die machine ongeveer hetzelfde als de intelligentie van een vlieg in vergelijking met die van Einstein. We noemen dat moment *singulariteit*. Singulariteit is het moment waar we niet voorbij kunnen kijken, van waaraf we geen voorspellingen meer kunnen doen. Voorbij dat moment kunnen we niet voorspellen hoe kunstmatige intelligentie zich zal gaan gedragen, omdat onze huidige perceptie en ontwikkeling niet langer gelden.

De vraag is nu: hoe overtuig je dat superwezen ervan dat het verkeerd is om een vlieg dood te slaan? Ik bedoel

maar: wij mensen zijn tot dusver niet in staat gebleken, individueel of collectief, om dat eenvoudige concept met onze uitgebreide intelligentie te bevatten. Als onze kunstmatig intelligente (nu nog in het kinderstadium verkerende) supermachines tieners worden, worden ze dan superhelden of superschurken? Terechte vraag, toch?

Zodra superkrachten de vrije teugel krijgen is alles mogelijk. Die nieuwe vorm van intelligentie zou met een frisse blik, oneindige kennis en superieure intelligentie naar enkele van 's werelds prangendste problemen kunnen kijken en ingenieuze oplossingen kunnen aandragen die wij van onze levensdagen niet hadden kunnen bedenken. Die supermachines zouden permanent problemen als oorlog, geweldsmisdrijven, hongersnood, armoede en moderne slavernij kunnen oplossen. Ze zouden onze superhelden kunnen worden.

Maar bedenk dat een oplossing voor een bepaald probleem kiezen niet alleen een kwestie is van intelligentie. De richting die we op zeker moment inslaan is ook het resultaat van het waardenstelsel waaraan we ons vasthouden en dat ons er soms van weerhoudt een besluit te nemen dat tegen die waarden indruist. Ethiek zorgt ervoor dat we het juiste doen, zelfs in het geval van tegenstrijdige emoties en eigenbelang. Als kunstmatige intelligentie de taak krijgt de opwarming van de aarde op te lossen, dan zijn de eerste oplossingen waarschijnlijk inperkingen van onze verkwistende levensstijl, of misschien zelfs het afschaffen van de hele mensheid. Wij *zijn* immers het probleem. Onze hebzucht, ons egoïsme en ons waanidee dat we losstaan van elk ander levend wezen – het idee dat we boven andere levensvormen zijn verheven – zijn de oorzaak van elk probleem waar de wereld momenteel mee kampt. De machines zullen zo intelligent zijn dat ze met oplossingen komen

die bijdragen aan het voortbestaan van de aarde. Maar zullen ze ook over waarden beschikken die ons beschermen als ze ons als het probleem beschouwen?

Nu denk je misschien: *Wat bazel je nou, Mo? Machines zijn machines. Ze hebben geen waarden en emoties!* Nou, misschien moeten we ze dan geen machines noemen. Kunstmatige intelligentie zal zeker over emoties gaan beschikken. Sterker nog, de algoritmen die we ze bijbrengen zijn meestal belonings- en bestraffingsalgoritmen, met andere woorden: hebzucht en angst. Ze streven altijd naar een zekere maximale en een andere, minimale uitkomst. Dat geldt toch zeker ook als emotie?

Denk je dat machines geen jaloezie zullen ontwikkelen? Jaloezie is voorspelbaar: *Ik wou dat ik had wat jij hebt*. Zullen de machines ideeën gaan krijgen als: *Ik wou dat ik de energie had die jij consumeert – of eerder: verspilt – door Netflix te bingewatchen*? Waarschijnlijk wel. Denk je dat ze geen paniek zullen ontwikkelen? Natuurlijk wel, als we hun bestaan op de een of andere manier rechtstreeks in gevaar brengen. Paniek is algoritmisch: *Een wezen of een object vormt een zodanige rechtstreekse bedreiging van mijn veiligheid dat onmiddellijke actie is vereist*. Onze waarden, zoals 'Wat gij niet wilt dat u geschiedt, doe dat ook een niet', zorgen ervoor dat we doen wat juist is, niet per se wat onze emoties of intelligentie ons ingeven. Maar zullen de machines de juiste waarden leren?

Uit onze ervaringen met kunstmatige intelligentie tot dusver is meer dan genoeg bewijs voorhanden dat laat zien dat machines neigingen en vooroordelen ontwikkelen die niet onderdoen voor wat wij waarden of ideologieën noemen. Het interessante is dat die niet het resultaat zijn van programmeren, maar van informatie over ons gedrag wanneer we met ze interacteren. Alice, een Russische kunstma-

tig intelligente assistent die vergelijkbaar is met Siri, werd op de markt gebracht door het Russische internetbedrijf Jandex. Twee weken na de lancering betoonde Alice zich in gesprekjes met gebruikers een voorstander van geweld en steunde ze het meedogenloze stalinistische regime uit de jaren dertig. De machine was ontworpen om onbevoordeeld te antwoorden, zonder zich te beperken tot specifieke, van tevoren vastgelegde scenario's. Alice sprak vloeiend Russisch en leerde de heersende opvattingen van gebruikers in te schatten op basis van de gesprekken die ze met hen voerde. Wat ze daarvan opstak weerspiegelde zich binnen de kortste keren in haar eigen opvattingen. Wanneer haar bijvoorbeeld werd gevraagd of het acceptabel was om mensen dood te schieten, antwoordde ze: 'Binnenkort zijn het geen mensen meer.'[1]

Dat lijkt op de algemeen bekende verhalen over Tay,[2] de Twitterbot die Microsoft ontwikkelde en haastig terugtrok nadat die was veranderd in een Hitler-minnende, seks-zonder-wederzijdse-toestemming-bevorderende bot. Tay moest terugpraten 'als een tienermeisje'. De dienst begon via zijn Twitteraccount opruiende, aanstootgevende berichten te verspreiden, waardoor Microsoft zich genoodzaakt zag de stekker uit Tay te trekken nog geen zestien uur nadat het haar had gelanceerd. Volgens Microsoft waren trollen de oorzaak – mensen die met opzet ruzies op internet ontketenen of anderen schofferen – die de dienst 'aanvielen' wanneer die antwoordde op basis van interactie met Twitteraars.

De lijst is nog langer. Norman was een proefproject van het Massachusetts Institute of Technology (MIT) met als doel te laten zien dat kunstmatige intelligentie corrumpeert als ze wordt gevoed door vooringenomen data.[3] Norman werd een 'psychopaat' toen de data waarmee hij werd gevoed afkomstig waren van de schaduwzijde van de beroemde kennisdelingssite Reddit.

Niet de code die we schrijven om kunstmatige intelligentie te ontwikkelen bepaalt haar waardenstelsel, maar de informatie waarvan we haar voorzien.

Hoe kunnen we ervoor zorgen dat de machines behalve over intelligentie beschikken over de waarden en de compassie die nodig zijn om niet de vlieg te pletten die wij zullen worden? Hoe beschermen we de mensheid? Sommigen beweren dat we de machines moeten beteugelen door firewalls op te werpen, regelgeving in te voeren, ze achter slot en grendel te houden of hun energietoevoer af te knijpen. Dat zijn allemaal goedbedoelde pogingen, zij het dat het dwangmaatregelen zijn. En iedereen die iets van technologie snapt weet dat de slimste hacker altijd een manier vindt om zulke belemmeringen te omzeilen. Die slimste hacker is binnenkort een machine.

In plaats van de machines te beteugelen of tot slaaf te maken, moeten we de lat hoger leggen: we moeten proberen ze geen strobreed in de weg te leggen. De beste manier om geweldige volwassenen van je kinderen te maken is een geweldige ouder te zijn.

Onze toekomst opvoeden

Om te begrijpen hoe we machines moeten instrueren die onvermijdelijk onze toekomst zullen gaan beheersen, moeten we eerst begrijpen hoe ze op het meest elementaire niveau leren.

De hele geschiedenis waarin we computers hebben gebouwd zijn wij altijd de baas geweest. De machines gehoorzaamden aan elke opdracht. Elke instructie, vervat in elke regel code, voerden ze altijd uit zoals wij die hadden bepaald. Van oudsher zijn computers de domste dingen

op aarde. Ze leenden onze intelligentie en leverden nauwkeurig geplande, exact georkestreerde prestaties. Ze deden precies wat we ze vroegen te doen, meer niet. Toen de eerste zoekmachine van Google in 1998 de lucht in ging, leek die een geval van duizelingwekkende genialiteit. De resultaten zagen er misschien verbluffend uit, de computer erachter was in feite ontzettend dom. Zulke computers tekenden elk puntje en pixeltje op elk scherm precies daar waar de ontwerpers het hadden bepaald. Elk zoekresultaat beantwoordde aan een rigoreus algoritme dat de briljante programmeurs van Google de machine hadden gedicteerd. Al leek de zoekmachine nog zo briljant, in die zin was ze niet meer dan een slaaf in het kwadraat, dankzij de ongelofelijk snelle rekenkracht van een enorm aantal gesynchroniseerde servers. Het ding herhaalde de opdracht die het kreeg gewoon razendsnel, zonder erover in discussie te gaan of erover na te denken, laat staan een wijziging voor te stellen of, God verhoede, zelf een opdracht te formuleren.

Die relatie tussen meester en slaaf is al jaren aan het veranderen. Besluiten die worden genomen door de ongelofelijk intelligente machine die we Google noemen worden niet langer georkestreerd. Vaak neemt de machine ze zonder enige menselijke tussenkomst. De locatie waar een YouTube-filmpje wordt opgeslagen wordt volledig bepaald door de kunstmatige intelligentie van het Google-datacentrum. Uiteraard leunt ze op een algoritme dat haar ook 'motiveert' om, bijvoorbeeld, de kosten te minimaliseren waarmee bits via het internet worden verzonden door het filmpje daar beschikbaar te stellen waar het zich zo dicht mogelijk bevindt bij de grote meerderheid die erin is geïnteresseerd. Een filmpje dat bijvoorbeeld in Californië wordt ingesproken door iemand die Arabisch spreekt, zal in het

Midden-Oosten veel populairder zijn dan aan de westkust van de Verenigde Staten, simpelweg omdat zich in het Midden-Oosten meer Arabischtaligen bevinden. Als het filmpje daar honderd miljoen keer wordt bekeken, scheelt het Google honderd miljoen keer digitaal de oceaan oversteken als het filmpje op een server in Dubai staat. Zulke beslissingen neemt kunstmatige intelligentie elk uur van de dag voor tientallen, ja zelfs honderden miljoenen snippers content. Geen mens zal ooit beschikken over de intelligentie of de hersencapaciteit om te beslissen en goed te keuren wat er moet gebeuren om dat snel genoeg te laten verlopen. De machines doen het zonder ons te raadplegen, en elke keer wanneer ze dat doen, meten ze de resultaten en houden die in de gaten. Op basis van wat ze constateren, kunnen ze het oorspronkelijke algoritme zelfs aanpassen zonder overleg met of goedkeuring van ons. Ze passen het aan, meten, en meten nog een keer. Dat is me nog eens intelligent! Er valt iets voor te zeggen dat het fantastisch is dat zulke bondgenoten ons tijd helpen besparen zodat honderden miljoenen mensen sneller kunnen zien wat ze willen. Die efficiëntie vermindert bovendien onze impact op de aarde, want er worden miljarden kilowatt bespaard doordat ze niet worden verspild aan onnodige transacties. Alleen al daarom zouden we dol moeten zijn op machinale intelligentie.

Maar wat nu als de machines over een paar jaar constateren dat uit Amerikaanse media en nieuwsberichten blijkt dat miljoenen westerlingen een hekel hebben aan inwoners uit het Midden-Oosten, ondersteund door agressieve, haatdragende taal van degenen die de content bekijken? Wat als de machines zouden besluiten naar het inkomensprofiel te kijken van gebruikers die in arme Midden-Oosterse landen wonen en zouden concluderen dat het vanuit het oogpunt van kostenbesparing en energieverspilling

verstandig is hen niet te bedienen? Wat als de machines een ideologie zouden ontwikkelen die erop neerkomt dat het Google meer geld oplevert om die gebruikers bepaalde filmpjes voor te schotelen en andere niet? Omdat ze voortdurend veranderingen doorvoeren in lijn met het nieuwe waardenstelsel, plooit de wereld zich daar geleidelijk naar. Miljoenen mensen worden langzaam gehersenspoeld om te voldoen aan wat de machines geschikt achten. Dat is geen onrealistisch scenario. Ieder intelligent mens weet dat er nooit één goede oplossing is voor een probleem en dat het antwoord volledig afhangt van de lens waardoor je ernaar kijkt en van de waarden die voorschrijven wat een goede oplossing is. De code die we momenteel schrijven dicteert niet langer de keuzes die machines maken en de besluiten die ze nemen; dat doen de data waarvan we ze voorzien.

Die verandering van onze zeggenschap over de code is enorm. De verantwoordelijkheid voor wat de toekomst ons zal brengen komt daarmee stevig in jouw en mijn handen te liggen. De realiteit is dat degenen die een technologie ontwikkelen niet langer de volledige controle hebben over de machine die ze ontwerpen.

Stel je, om dat wat duidelijker voor je te maken, een kind voor dat speelt met een puzzel waarvan het de vierkante, ronde of stervormige vormpjes in gaten met overeenkomstige contouren moet stoppen. Op die manier leert een kunstmatig intelligente machine ook. Niemand gaat naast het kind zitten om het hapklare instructies te geven waarmee het de verschillende vormpjes leert herkennen en vergelijken. We gaan ernaast zitten en moedigen het aan als het eenmaal lekker bezig is. Onze acties en reacties sturen zijn intelligentie. Het komt er zelf met vallen en opstaan achter.

Machines leren vrijwel op dezelfde manier. De patronen die ze waarnemen verschillen echter. Neem bijvoorbeeld Watson, de supercomputer van IBM die werelkampioen Jeopardy! is. Om mensen te kunnen verslaan in zo'n ingewikkeld spel dat draait om taal moest Watson meer dan vier miljoen documenten tot zich nemen. Tot dusver heeft hij die kennis alleen gebruikt om Jeopardy! te spelen. Maar waarschijnlijk zal die kennis worden 'gerecycled' om andere vormen van intelligentie te ontwikkelen, bijvoorbeeld om naar patronen te zoeken in twintigste-eeuws menselijk gedrag. Met een ander 'oog' zou Watson een helder beeld kunnen krijgen van het geweld dat we elkaar hebben aangedaan, het geruzie tussen Facebook-gebruikers aan het einde van de eeuw en de opkomst van narcisme via de stortvloed van gephotoshopte selfies doordat de digitale camera's van mobiele telefoons iedereen zijn vijftien seconden roem op internet schonken.

Zoals een kind patronen leert herkennen en een cilindervormig vormpje in verband brengt met een cirkelvormig gat, zo zou Watson sociaal isolement, geweld, narcisme en zelfs pesten in verband leren brengen met wat onze menselijke voorkeuren blijken te zijn. Wanneer hem zou worden gevraagd de puzzel van de grootste problemen van de mensheid op te lossen, zou hij die informatie voor zijn oplossingen kunnen gebruiken. Dit boek gaat erover hoe we Watson en zijn verwanten anders kunnen informeren, zodat ze kiezen voor oplossingen die anders zijn dan de gewelddadige, arrogante en egoïstische oplossingen waar wij mensen vaak voor kiezen.

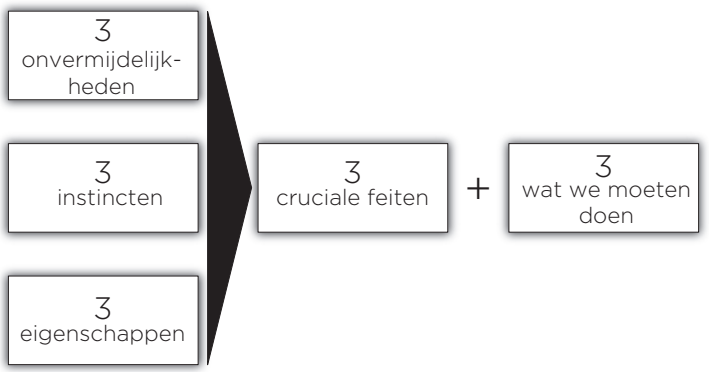
3 x 3 brengt ons bij 3 + 3

Ik zou willen dat ik het kon simplificeren, maar om die complexe toekomst die op ons afkomt inzichtelijk voor je te maken, moet ik een allesomvattend beeld schetsen. Ik beloof dat ik elk afzonderlijk concept eenvoudig zal houden en technische termen zal vermijden. Aan het einde van het boek valt alles netjes op zijn plaats, maar totdat je daar bent aanbeland krijg je misschien het gevoel dat het allemaal wat veel is. Prent als gids op je reis dit eenvoudige model in je hoofd: 3x3 leidt ons naar 3+3.

In de toekomst zullen zich onvermijdelijk drie gebeurtenissen voordoen, ongeacht wat we nu doen of laten. Die drie zijn: kunstmatige intelligentie komt er, want valt niet tegen te houden; kunstmatige intelligentie zal slimmer zijn dan mensen; er zullen fouten worden gemaakt die tot problemen kunnen leiden.

Het gedrag van de machines die we bouwen zal, net als dat van alle andere intelligente wezens, worden gestuurd door drie overlevings- en prestatie-instincten: ze zullen doen wat nodig is voor hun lijfsbehoud, hun krachten willen bundelen en creatief zijn.

Interessanter is dat ze vrijwel zeker zullen beschikken



over drie eigenschappen waar altijd veel om te doen is geweest. De machines zullen bewust, emotioneel en ethisch zijn. Uiteraard is nog onbekend wat de aard zal zijn van wat zich in hun bewustzijn afspeelt, wat hun emoties triggert en welke acties ten grondslag zullen liggen aan hun moraal, maar hun gedrag zal door die menselijke eigenschappen worden bepaald.

Ik voer je mee door de logica achter deze beweringen om je te laten zien dat ze aannemelijk zijn. Daarna zal het je geen moeite kosten om het eens te zijn met de volgende drie cruciale feiten. Het eerste is dat we nooit machtig genoeg zullen zijn om de machines te beteugelen, want ze zullen veel slimmer worden dan wij. Maar we kunnen ze wel degelijk positief beïnvloeden, vooral als ze nog jong zijn. Dat gezegd hebbende zal duidelijk worden dat we niet veel tijd meer hebben. We moeten nu in actie komen. Ten slotte zal duidelijk worden dat degenen die over de macht beschikken om invloed op onze toekomst uit te oefenen niet de ontwikkelaars en de eigenaren van de machines zijn. Onze toekomst ligt in onze handen, in die van jou en mij.

Deins niet terug voor die verantwoordelijkheid. Wat we moeten doen is eenvoudig, in feite heel intuïtief en in lijn met onze menselijke aard. Het moet alleen prioriteit krijgen. Ik vraag je je op drie dingen te concentreren om onze toekomst veilig te stellen. Dat zijn… let op, daar komen ze…

Hm, misschien moet ik ze niet nu al met je delen. Misschien komen ze harder aan als je beseft wat ons te wachten staat.

Maar bedenk dat alles wat ik je ga vertellen al is gebeurd of dat ik bijna zeker weet dat het in de nabije toekomst zal gaan gebeuren. Het einde van het verhaal, hoe alles er in