

Inhoud

Proloog	13
1. Inleiding: De ware mens	15
2. Verificatie	29
3. De reizende ziel	52
4. Locatiespecificiteit versus pure techniek	90
5. Buiten de boekjes	116
6. Doel	151
7. Binnenvallen	171
8. De slechtste getuige ter wereld	197
9. Niet intact blijven	221
10. Grote verrassing	247
Conclusie: De ware mens	291
Epiloog: De nooit bezongen schoonheid van de glazenkast	299
Dankbetuigingen	304
Noten	307

1. Inleiding

De ware mens

Ik word 7500 kilometer van huis wakker in een hotelkamer zonder douche; voor de eerste keer in vijftien jaar neem ik een bad. Ik eet (heel traditioneel) een paar twijfelachtig ogende tomaten, wat gebakken bonen en vier halve geroosterde witte boterhammen die op een klein metalen rekje worden gepresenteerd, verticaal, alsof het boeken zijn. Daarna stap ik naar buiten, de zilte lucht in, en loop langs de kust van het land waar mijn taal is uitgevonden, hoewel ik tamelijk veel borden waar ik langs loop niet begrijp: LET AGREED staat er bijvoorbeeld nadrukkelijk met grote letters, en dat zegt me niets.

Ik sta stil en kijk een ogenblik naar de zee terwijl ik de tekst in mijn hoofd probeer te analyseren. Normaal vind ik dit soort taalkundige curiositeiten en culturele kloven intrigerend, maar vandaag maak ik me er vooral zorgen over. De volgende twee uur zal ik achter een computer zitten en een reeks chats van vijf minuten per stuk hebben met diverse onbekenden. Aan de andere kant van de chats zullen een psycholoog, een taalkundige, een computerwetenschapper en de host van een populaire Britse technologieshow zitten. Gezamenlijk vormen ze een panel dat moet oordelen, en het doel van deze conversaties is een van de vreemdste dingen die me ooit zijn gevraagd.

Ik moet ze ervan overtuigen dat ik menselijk ben.

Gelukkig ben ik menselijk; jammer genoeg is het niet duidelijk of dat wel zal helpen.

De Turingtest

Elk jaar komt de gemeenschap voor artificiële intelligentie (AI) vol hooggespannen verwachtingen bij elkaar voor het meest controversiële evenement van het vakgebied: een competitie die de Turingtest wordt genoemd. De test is genoemd naar de Britse wiskundige Alan Turing, een van de grondleggers van de computerwetenschappen, die in 1950 probeerde een van de eerste vragen in het vakgebied te beantwoorden: kunnen machines denken? Anders gezegd: zal het ooit mogelijk zijn om een computer te construeren die zo geavanceerd is dat we werkelijk kunnen zeggen dat hij denkt, intelligent is, een geest heeft? En als we dat ooit op een dag zouden kunnen, hoe zouden we dat dan weten?

Turing stelde voor om, in plaats van puur theoretisch over deze vraag te discussiëren, een experiment uit te voeren. Een panel van juryleden stelt via een computerscherm vragen aan twee onzichtbare correspondenten, de ene een menselijke medeplichtige, de andere een computerprogramma, en probeert te onderscheiden wie wie is. Er zijn geen beperkingen aan wat er mag worden gezegd; het gesprek kan variëren van koetjes en kalfjes tot aan de feiten van de wereld (bijvoorbeeld hoeveel poten mieren hebben, in welk land Parijs ligt enzovoort), tot aan roddels over beroemdheden en zwaarwichtige filosofie; kortom, het hele gamma van de menselijke conversatie. Turing voorspelde dat tegen het jaar 2000 computers 30 procent van de juryleden na vijf minuten conversatie voor de mal zouden kunnen houden en dat we daarom ‘zouden kunnen zeggen dat machines denken zonder dat iemand ons tegensprekt’.

Turings voorspelling is niet uitgekomen, maar tijdens de test van 2008 in Reading (Engeland), kwam het beste programma maar één stem tekort. De test van 2009 zou wel eens beslissend kunnen zijn.

En ik neem daaraan deel als een van de vier menselijke medeplichtigen die het opnemen tegen de beste AI-programma's. In een aantal ronden word ik, net als de andere medeplichtigen, ge-

koppeld aan een AI-programma en een jurylid, en moet ik de laatste ervan overtuigen dat ik daadwerkelijk menselijk ben.

Het jurylid praat vijf minuten met een van ons, dan vijf minuten met de ander, en heeft vervolgens tien minuten om na te denken en te kiezen wie van ons volgens hem menselijk is. Juryleden noteren ook op een glijdende schaal hoeveel vertrouwen ze in dit oordeel hebben; dit wordt deels gebruikt als beslissende maatstaf. Elk jaar ontvangt het programma dat de meeste stemmen en het meeste vertrouwen van de juryleden krijgt (ongeacht of het voor de Turingtest slaagt door 30 procent van de juryleden voor de mal te houden), de titel 'Most Human Computer' (menselijkste computer). Dit is de titel waar alle onderzoeksteams op uit zijn, de titel met de geldprijs, waar de organisatoren en het publiek het meest in geïnteresseerd zijn. Maar intrigerend genoeg is er ook nog een andere titel, namelijk die voor de medeplichtige die de meeste stemmen en het meeste vertrouwen van de juryleden krijgt: de prijs voor de 'Most Human Human' (de menselijkste mens, of de ware mens).

Een van de eerste winnaars, in 1994, was Charles Platt, columnist van *Wired*. Hoe kreeg hij het voor elkaar? Door 'humeurig, prikkelbaar en onhebbelijk te zijn', zegt hij, en dat lijkt me niet alleen geestig en somber, maar op een dieper niveau ook een oproep tot de strijd: hoe zijn wij eigenlijk de menselijkste, meest ware mensen die we kunnen zijn, niet alleen in de beperkende omstandigheden van een test, maar in het echte leven?

Meedoen aan de samenzwering

De sponsor en organisator van de Turingtest (deze specifieke gelegenheid staat bekend als de Loebnerprijs) is een kleurrijke en ietwat zonderlinge figuur: Hugh Loebner, de baron van de plastic oproldiscodansvloer. Als hem wordt gevraagd wat zijn motieven zijn om deze jaarlijkse Turingtest te ondersteunen en te organiseren, noemt Loebner nota bene 'luiheid': zijn utopische toekomst

is er kennelijk een waarin het werkeloosheidscijfer bijna 100 procent is en bijna alle menselijke inspanningen en nijverheid zijn uitbesteed aan intelligente machines. Ik moet zeggen dat deze visie van de toekomst mij vooral wanhopig voorkomt; ik heb zelf heel andere ideeën over hoe een door AI bevolkte wereld eruit zou zien, en ik heb heel andere redenen om mee te doen met de test. Maar in elk geval is de cruciale vraag de centrale vraag hoe computers ons besef van een ‘zelf’ hervormen, en wat de uitbreidingen van dat proces zullen zijn.

Ik wist niet helemaal zeker hoe ik een medeplichtige moest worden en begon bovenaan, door te proberen met Hugh Loebner zelf in contact te komen. Ik had zijn website snel gevonden, waarop ik temidden van een tamelijk ondoordringbaar mengsel van materiaal over paaltjes voor het in bedwang houden van een menigte,¹ prostitutieactivisme² en een schandaal over de samenstelling van olympische medailles,³ informatie kon vinden over zijn gelijknamige prijs en zijn e-mailadres. Als antwoord verwees hij me naar Philip Jackson, een jonge docent aan de University of Surrey, die de leiding heeft over de logistiek van de strijd om de Loebnerprijs van dit jaar in Brighton, onder de auspiciën van de InterSpeech Conference van 2009 over spraak- en communicatiewetenschap.

Ik kon via Skype in contact komen met professor Jackson, een jonge, slimme knaap met dat typische vermoeide enthousiasme dat kenmerkend is voor een overwerkte maar verse academicus

-
- 1 Het schijnt dat de dwangpaaltjes voor menigten onlangs de draagbare discodansvloeren hebben vervangen als het vlaggenschip van Loebners bedrijf, Crown Industries, de voornaamste sponsor van de Loebnerprijs.
 - 2 Ik ben vast niet de enige die het ironisch vindt dat een man die zich wijdt aan de vooruitgang in de interactie met *artificiële* entiteiten erin berust (zoals hij openlijk heeft besproken in de *New York Times* en in diverse televisietalkshows) om, al of niet met tegenzin, te betalen voor *menselijke* intimiteit.
 - 3 Kennelijk zijn de ‘gouden’ medailles eigenlijk zilveren medailles die *in goud zijn gedoopt*; dat is wel wat bizar, maar het heeft bij Loebner blijkbaar al tien jaar verontwaardiging opgeleverd die zich heeft geuit in de vorm van negatieve publiciteit en speeches, en een nieuwsbrief die *Pants on Fire News* heet.

die op een vaste aanstelling uit is. Daarbij zijn charmante Britse trekjes, zoals ‘skeletal’ zo uitspreken dat het rijmt op ‘a beetle’: ik mocht hem direct. Zijn werk richt zich vooral op spraakanalyse, in tegenstelling tot de eigenlijke artificiële intelligentie; het was moeilijk te zeggen of hij echt opgewonden was dat hij de Loebnerprijs van dit jaar mocht organiseren, of dat het gewoon een klus was voor de laagste ambtenaar zonder vaste aanstelling bij Inter-Speech; misschien wel allebei.

Hij vroeg wat ik deed en ik legde uit dat ik non-fictie schrijf over wetenschap en filosofie, vooral over de manier waarop wetenschap en filosofie het dagelijks leven doorkruisen, en dat ik gefascineerd ben door de idee van de Turingtest en de ‘ware mens’. Het is bijvoorbeeld een romantische gedachte om samen het menselijk ras te verdedigen, zoiets als Gary Kasparov versus Deep Blue, en binnenkort Ken Jennings van *Jeopardy!* versus het nieuwste IBM-systeem genaamd Watson. (Er komen ook andere, meer *Terminator 2*- en *The Matrix*-achtige fantasieën bovendrijven, hoewel de Turingtest aanzienlijk minder machinegeweren belooft.) Toen ik las dat de machines de test van 2008 op één stem na hadden gehaald, en besepte dat 2009 het jaar zou kunnen worden waarin ze die drempel eindelijk zouden passeren, kwam er schijnbaar uit het niets een vastberaden stem bij me op: over mijn lijk.

Maar wat belangrijker is, de test roept een aantal opwindende en verontrustende vragen op over die kruising van computerwetenschap, cognitiewetenschap, filosofie en het dagelijks leven. Ik heb elk van deze gebieden bestudeerd en erover geschreven, onderzoek gepubliceerd over cognitiewetenschap dat door vakgenoten is gerecenseerd, en als zodanig vind ik de Turingtest met name fascinerend vanwege de manier waarop die al deze gebieden bij elkaar brengt. Tijdens het chatten vertelde ik aan Jackson dat ik dacht dat ik misschien een unieke bijdrage aan de Loebnerprijs kon leveren, zowel in mijn daadwerkelijke kwaliteit van medeplichtige, als in het verslaan van deze ervaring voor een groot publiek, samen met de bredere vragen en problemen die de test oproept; dat zou naar mijn mening een boeiende en belangrijke

conversatie in de bredere openbare cultuur teweeg kunnen brengen. Het was niet zo moeilijk om hem daarmee te laten instemmen en al gauw stond mijn naam op de medeplichtigenlijst.

Nadat hij me een paar praktische dingen over de logistiek van de competitie had verteld, gaf hij me het advies dat ik al van eerdere medeplichtigen had gehoord: ‘Meer hoeft je eigenlijk niet te weten. Je bent menselijk, dus wees gewoon jezelf.’

‘Wees gewoon jezelf’: dit is in feite het motto van de medeplichtigen geweest sinds de eerste Loebnerprijs in 1991, maar dat is wat mij betreft een wat naïef groot vertrouwen in menselijke instincten, of op zijn ergst een manier om de strijd van tevoren al te beslissen. De programma’s voor artificiële intelligentie waar we het tegen opnemen, zijn vaak het resultaat van tientallen jaren werk; goed, dat zijn wij ook. Maar de onderzoeksteams voor artificiële intelligentie hebben enorme databases met testuitvoeringen van hun programma’s en ze hebben statistische analyses op die archieven uitgevoerd: ze weten hoe ze de conversatie handig weg moeten leiden van hun tekortkomingen in de richting van hun sterke punten, welke gespreksroutes leiden tot diepzinnige uitwisselingen en welke mislukken; de straatvechtersinstincten van de gemiddelde medeplichtige zijn vast niet zo goed. Dit is een eigenaardig en heel interessant punt en er is ruimschoots bewijs voor; in onze samenleving is er voortdurend vraag naar coaches voor gesprekken, spreken in het openbaar en dating. In de transcripten van de test van 2008 zien we dat de juryleden zich bepaald verontschuldigend gedragen tegenover de menselijke medeplichtigen omdat ze geen betere conversatie kunnen leveren: ‘Het spijt me voor de [medeplichtigen], ze zullen zich wel vervelen met dat praten over het weer,’ zegt er een, en een ander merkt deemoedig op: ‘Sorry dat ik zo banaal ben’, terwijl de computer in het andere venster het jurylid blijkbaar helemaal inpakt, want die schrijft al snel de ene *LOI* na de andere: *P*. Dat kunnen we beter.

Ik moet dan ook zeggen dat het vanaf het begin mijn intentie was om zo ongehoorzaam mogelijk te zijn aan het advies van de organisatoren om ‘gewoon in september naar Brighton te komen

en mezelf te zijn', dat ik de maanden voor de test doorbracht met het verzamelen van zo veel mogelijk informatie, voorbereiden en het opdoen van ervaring, en dat ik naar Brighton kwam met het vaste voornemen om alles uit mezelf te halen.

Eigenlijk zou dit helemaal niet zo vreemd moeten zijn, natuurlijk; we oefenen en we bereiden ons voor op tenniscompetities, spellingwedstrijden en schooltoetsen en dergelijke. Maar omdat de Turingtest bedoeld is om te beoordelen hoe menselijk ik ben, lijkt de implicatie dat menselijk zijn (en jezelf zijn) meer is dan alleen maar komen opdagen. Ik houd vol dat dat inderdaad zo is. Wat dat 'meer' precies inhoudt, vormt het hoofdthema van dit boek, en de antwoorden die we onderweg vinden, zullen van toepassing zijn op heel wat meer dan alleen de Turingtest.

Voor Ivana vallen

En nu een eigenaardig en tamelijk ironisch verhaal met een waarschuwing: dr. Robert Epstein, psycholoog aan de UCSB, redacteur van de wetenschappelijke publicatie *Parsing the Turing Test* en met Hugh Loebner medeoprichter van de Loebnerprijs, meldde zich in de winter van 2007 aan bij een onlinedatingservice. Hij begon lange brieven te schrijven naar een Russische vrouw die Ivana heette en die lange brieven terugschreef, over haar familie, haar dagelijks leven en haar sterker wordende gevoelens voor Epstein. Maar uiteindelijk voelde het toch niet helemaal goed. En om een lang verhaal kort te maken: Epstein beseft ten slotte dat hij meer dan vier maanden lange liefdesbrieven had uitgewisseld met (u raadt het al) een computerprogramma. Arme knul! Alsof het nog niet genoeg was dat webschurken elke dag zijn e-mailpostvak vol spam gooiden – moesten ze dat nu ook al met zijn hart doen?

Aan de ene kant zou ik simpelweg achterover willen leunen en hem willen uitlachen; hij heeft nota bene de Loebnerprijs helpen oprichten! Wat een sukkel! Aan de andere kant heb ik ook sympathie voor hem: de onvermijdelijke spam in de eenentwintigste

eeuw verstoopt niet alleen de postvakken en de bandbreedte van de wereld (zo'n 97 procent van alle e-mailberichten is spam⁴; dit gaat over tientallen miljarden per dag. Je zou letterlijk een klein land⁵ kunnen voeden met de elektriciteit die nodig is voor het verwerken van de dagelijkse spam over de hele wereld), maar het doet ook nog iets veel ergers: het holt ons vertrouwen uit. Ik vind het een ramp dat ik, wanneer ik berichten van mijn vrienden krijg, minstens de eerste paar zinnen energie moet besteden om te beslissen of die berichten echt van mijn vrienden komen. In de eenentwintigste eeuw gaan we door het digitale leven met onze verdediging in stelling gebracht. Alle communicatie is een Turingtest. Alle communicatie is verdacht.

Dat is de pessimistische versie, en hier is de optimistische. Ik wed dat Epstein een lesje geleerd heeft, en ik wed ook dat die les een stuk ingewikkelder en subtieler was dan 'het was een slecht idee om te proberen een onlinerelatie op te starten met iemand uit Nizjny Novgorod'. Ik zou graag denken dat hij zich ernstig heeft afgevraagd waarom het vier maanden duurde voordat hij zich realiseerde dat er geen echte uitwisseling plaatsvond tussen hem en 'Ivana', en dat hij zoiets voortaan sneller door zal hebben. En dat zijn volgende vriendin, die hopelijk niet alleen een echte homo sapiens is maar ook wat dichterbij woont dan elf tijdzones uit de buurt, in zekere zin 'Ivana' dankbaar mag zijn.

De illegitimiteit van het overdrachtelijke

Toen Claude Shannon in de jaren veertig Betty ontmoette bij Bell Labs, was ze inderdaad een 'computer'. Dit klinkt ons misschien vreemd in de oren, maar laat ik u zeggen dat dat in die tijd niemand vreemd voorkwam. Ook hun collega's niet: bij Bell Labs

4 Volgens een citaat van Microsoft op de BBC op 8 april 2009: <http://news.bbc.co.uk/2/hi/technology/7988579.stm>.

5 Bijvoorbeeld Ierland.

vonden ze de romance volkomen normaal en zelfs voor de hand liggend. Engineers en computers flirtten continu met elkaar.

Het was Alan Turings artikel uit 1950, ‘Computing Machinery and Intelligence’, waarmee het vakgebied van AI werd gestart zoals we het nu kennen, en waarmee de conversatie en de controverse over de Turingtest (of het ‘imitatiespel’, zoals Turing het aanvankelijk noemde) werden gestart die tot op heden nog voortduren; maar moderne ‘computers’ zijn heel wat anders dan de ‘computers’ in de tijd van Turing. In het begin van de twintigste eeuw, voordat een ‘computer’ een digitaal verwerkingsapparaat was dat we in onze levens van de eenentwintigste eeuw zo vaak aantreffen (in onze kantoren, huizen, auto’s en ook steeds vaker in onze zakken), was het iets anders: een functieomschrijving.

Vanaf het midden van de achttiende eeuw waren ‘computers’, vaak vrouwen, in dienst bij bedrijven, technische firma’s en universiteiten om berekeningen en numerieke analyses uit te voeren, soms met de hulp van een basale rekenmachine. Deze oorspronkelijk menselijke computers zaten achter de berekeningen voor alles, vanaf de eerste nauwkeurige voorspellingen van de terugkeer van de komeet van Halley (een vroeg bewijs van Newtons zwaartekrachttheorie, die daarvoor alleen was gecontroleerd aan de hand van planeetbanen) tot aan het Manhattanproject, waarin de Nobelprijswinnaar en fysicus Richard Feynman de leiding had over een groep menselijke computers bij Los Alamos.

Je staat versteld als je een paar van de eerste artikelen over computerwetenschap leest en ziet hoe de auteurs hun best doen om voor het eerst uit te leggen wat deze nieuwe dingen nu eigenlijk waren. In Turings artikel wordt de ‘digitale computer’, waar nog nooit iemand van heeft gehoord, beschreven door middel van analogieën met een menselijke computer: ‘Het idee achter “digitale computers” kan worden uitgelegd door te stellen dat deze machines zijn bedoeld om alle bewerkingen uit te voeren die door een menselijke computer kunnen worden verricht.’ We weten natuurlijk dat in latere decennia de aanhalingstekens zijn verplaatst, en dat nu ‘digitale computer’ niet alleen de standaardterm is, maar

ook de letterlijke. En het is de ‘menselijke computer’ die is verwezen naar de illegitimiteit van het overdrachtelijke. In het midden van de twintigste eeuw was een nieuw, flitsend mathematisch snuffje ‘net een computer’. In de eenentwintigste eeuw is een menselijk wiskundewonder ‘net een computer’. Een eigenaardige wending: we lijken op dat ding dat vroeger op ons leek. We imiteren onze vroegere imitators, een van de vreemde omkeringen van het lot in de lange sage van de menselijke uitzonderlijkheid.

De Zin

Daniel Gilbert, psycholoog aan Harvard, zegt dat iedere psycholoog op een bepaald moment in zijn of haar loopbaan een versie moet schrijven van ‘De Zin’. De Zin luidt als volgt: ‘De mens is het enige dier dat...’ Het lijkt er inderdaad op dat filosofen, psychologen en wetenschappers deze zin hebben geschreven en herschreven sinds het begin van de geschreven geschiedenis. We zouden kunnen zeggen dat het verhaal van het menselijke zelfbewustzijn het verhaal is van mislukte, ontmaskerde versies van De Zin. Alleen zijn het nu niet uitsluitend de dieren waar we ons zorgen over maken.

Ooit dachten we dat mensen uniek waren omdat ze een taal met een syntaxis hadden, maar dat is niet waar;⁶ we dachten ooit dat mensen uniek waren omdat ze gereedschappen gebruikten, maar dat is niet waar;⁷ we dachten ooit dat mensen uniek waren

6 Michael Gazzaniga citeert in *Human Sue Savage-Rumbaugh*, primatoloog van de Georgia State University: ‘Eerst zeiden de taalkundigen dat we onze dieren tekens symbolisch moesten leren gebruiken om te kunnen zeggen dat ze een taal leerden. Goed, dat deden we, en toen zeiden ze: ‘Nee, dat is geen taal, want je hebt geen syntaxis.’ Daarom bewezen we dat onze apen combinaties van tekens konden produceren, maar de taalkundigen zeiden dat dat niet genoeg syntaxis was, of niet de goede syntaxis. Ze zullen nooit vinden dat we genoeg hebben gedaan.’

7 In 2009 werd bijvoorbeeld ontdekt dat octopussen kokosschalen konden gebruiken als ‘harnas’. De samenvatting van het artikel waarin dit nieuws stond, vertelt hoe onze vermeende uitzonderlijkheid steeds verder wordt uitgehoud:

omdat ze konden rekenen, maar nu kunnen we ons nauwelijks voorstellen dat wij zouden kunnen wat onze calculators kunnen.

Er zijn verschillende aspecten aan de evolutie van De Zin. We kunnen bijvoorbeeld bekijken hoe de formuleringen van diverse ontwikkelingen (zowel in onze kennis van de wereld als in onze technische vermogens) in de loop van de geschiedenis zijn veranderd. Van daaruit kunnen we bekijken hoe deze verschillende theorieën het identiteitsgevoel van de mensheid hebben vormgegeven. Zijn kunstenaars bijvoorbeeld waardevoller voor ons geworden dan toen we nog niet wisten hoe moeilijk kunst voor computers is?

Ten slotte kunnen we ons afvragen: is het terecht dat onze definitie van onze unieke positie in zekere zin een reactie is op de technologische vooruitgang? En waarom hebben we het eigenlijk zo hard nodig om ons uniek te voelen?

‘Het lijkt soms wel,’ zegt Douglas Hofstadter, ‘alsof elke nieuwe stap in de richting van artificiële intelligentie niet zozeer iets produceert dat volgens iedereen echte intelligentie is, maar eerder duidelijk maakt wat echte intelligentie niet is.’ Dit lijkt misschien op het eerste gezicht een troostend standpunt, dat onze unieke claim op het denken intact laat; maar het lijkt ook ongemakkelijk veel op een geleidelijke terugtrekking, ongeveer zoals een middeleeuws leger dat zich van het kasteel terugtrekt op de toren. Maar deze terugtrekking kan niet oneindig lang doorgaan. Stel dat van alles wat we als denken beschouwen, blijkt dat het dat niet is: wat is ‘denken’ dan? Denken lijkt te worden gereduceerd tot een epifenomeen, een soort afvalproduct van het brein; of erger nog, tot een illusie.

‘Gedrag waarbij gereedschappen worden gebruikt, werd oorspronkelijk beschouwd als een definiërend kenmerk van onze soort, maar later is gebleken dat ook andere primaten en ook zoogdieren en vogels dit gedrag vertonen. Van ongewervelden was echter nooit eerder gemeld dat ze objecten verzamelen om die later te gebruiken. We hebben herhaaldelijk gezien hoe octopussen die in zachte sedimenten leven, halve kokosnootschalen meedragen en die alleen tot een beschutting samenvoegen wanneer dat nodig is.’

Waar is de toren van ons zelf?

Het verhaal van de eenentwintigste eeuw zal deels het verhaal zijn van het steeds opnieuw trekken van deze scheidslijnen, het verhaal van homo sapiens die een claim probeert te leggen op verschuivende grond, geflankeerd door aan de ene kant het dier en aan de andere kant de machine, ingeklemd tussen bieflappen en bits.

En hier is een cruciale vraag: is deze terugtrekking nu goed of slecht? Ontneemt bijvoorbeeld het feit dat computers zo goed zijn in wiskunde, ons een deel van onze menselijke activiteiten, of bevrijdt het ons juist, in die zin dat we geen niet-menselijke activiteit meer hoeven te verrichten en een menselijker leven kunnen leiden? Deze laatste opvatting lijkt misschien aantrekkelijker, maar dat wordt minder als we ons indenken dat er in de toekomst een moment kan komen dat het aantal 'menselijke activiteiten' dat we nog kunnen doen, onplezierig klein is geworden. En wat dan?

De Turingtest omkeren

Er bestaan geen bredere filosofische implicaties. [...] Het houdt nergens verband mee en verheldert niets.

– NOAM CHOMSKY in een e-mail aan de auteur

Alan Turing stelde zijn test voor als een manier om de vooruitgang van de technologie te meten, maar hij biedt ons net zo goed een manier om onze eigen vooruitgang te meten. Oxfordfilosoof John Lucas zegt bijvoorbeeld dat, wanneer we niet in staat zullen blijken om te verhinderen dat de machines slagen voor de Turingtest, dat niet komt 'doordat machines zo intelligent zijn, maar doordat mensen, in elk geval een groot aantal van hen, zo inflexibel zijn'.

Het komt hierop neer: de Turingtest is niet alleen een technologische toets, en stelt niet alleen filosofische, biologische en ethische vragen; de test gaat bovenal over de handeling van het communiceren. De diepzinnigste vragen rondom de test zie ik als

praktische vragen. Hoe kunnen we binnen de grenzen van taal en tijd betekenisvol, of zo betekenisvol mogelijk, met elkaar communiceren? Hoe werkt empathie? Wat is het proces waardoor iemand in ons leven komt en iets voor ons gaat betekenen? Voor mij zijn dit de centrale vragen van de test, en de centrale vragen van het mens-zijn.

Het fascinerende aan het bestuderen van de programma's die goed hebben gescoord bij de Turingtest is onder andere dat dit een (eerlijk gezegd ontvullend) onderzoek is naar de manier waarop conversatie kan werken wanneer er geen enkele emotionele intimiteit aanwezig is. De transcripten van oudere Turingtests bekijken is een soort rondleiding in de verschillende manieren waarop we protesteren, vragen ontwijken, de stemming verluchten, van onderwerp veranderen, afleidingsmanoeuvres uitvoeren, de tijd doden: alles wat bij de Turingtest niet voor echte conversatie zou moeten doorgaan, zou waarschijnlijk ook niet mogen doorgaan voor echte menselijke conversatie.

Er zijn diverse boeken geschreven over de technische kant van de Turingtest, bijvoorbeeld over het slim ontwerpen van Turingtestprogramma's, ook wel 'chatterbots', 'chatbots' of gewoon 'bots' genoemd. In feite gaan alle praktische teksten over de Turingtest over hoe we goede bots moeten maken, en een klein deeltje gaat over hoe we goede juryleden kunnen zijn. Maar je leest nergens hoe je een goede medeplichtige moet zijn. Ik vind dat vreemd, want het is volgens mij juist aan de kant van de medeplichtige waar de inzet het hoogst is en waar de antwoorden de meest verstrekkende gevolgen hebben.

Ken uw vijand beter dan uzelf, vertelt Sun Tzu ons in *De kunst van het oorlogvoeren*. In het geval van de Turingtest betekent het kennen van onze vijand in werkelijkheid dat we onszelf moeten kennen. Daarom zullen we wel kijken naar hoe sommige van deze bots zijn geconstrueerd, en naar een paar van de basisprincipes en de belangrijkste resultaten in de theoretische computerwetenschap, maar altijd met de focus op de menselijke kant van de vergelijking.

In zekere zin is dit een boek over artificiële intelligentie, het verhaal van de geschiedenis ervan, en van mijn bescheiden persoonlijke betrokkenheid bij die geschiedenis. Maar in de kern is dit een boek over leven.

We kunnen computers, die een steeds centralere rol in ons leven spelen, beschouwen als verwoesters: een kracht als Skynet in *Terminator*, of de Matrix in *The Matrix*, die op onze vernietiging uit is, net zoals wij op hun vernietiging uit zouden moeten zijn. Maar ik geef om diverse redenen de voorkeur aan het idee van rivalen, die duidelijk willen winnen en die weten dat het hoofddoel van de competitie draait om het verhogen van het spelniveau. Alle rivalen zijn symbioten. Ze hebben elkaar nodig. Ze houden elkaar eerlijk. Ze maken elkaar beter. Het verhaal van de vooruitgang hoeft geen ontmenselijkend of ontmoedigend verhaal te zijn. Integendeel, zoals u zult zien.

Tijdens de maanden voor de test deed ik alles wat ik kon om me voor te bereiden. Ik deed onderzoek en sprak met deskundigen op allerlei gebieden die iets te maken hadden met de centrale vragen: (a) hoe ik in Brighton de ‘menselijkste’ vertoning ten beste kon geven, en (b) wat het nu eigenlijk inhoudt om menselijk te zijn. Ik interviewde onder meer taalkundigen, informatietheoretici, psychologen, advocaten en filosofen; deze gesprekken bezorgden me zowel praktische adviezen voor de competitie als kansen om te bekijken hoe de Turingtest (met zijn bijbehorende vragen naar menselijkheid) invloed heeft op, en wordt beïnvloed door uiteenlopende gebieden als werk, school, schaken, dating, videogames, psychiatrie en het recht.

De uiteindelijke test hield voor mij in dat ik in Brighton een zo uniek menselijk mogelijke demonstratie zou geven om me met succes te verdedigen tegen de machines die de test ondergingen, en dat ik die felbegeerde, zij het bizarre prijs voor de ware mens mee naar huis zou kunnen nemen. Maar de ultieme vraag was natuurlijk wat het betekent om menselijk te zijn: wat de Turingtest ons over onszelf kan leren.