# Excerpt

## A. M. Turing, Computing Machinery and Intelligence

G.J.E. Rutten

In genoemd artikel gaat Turing in op de vraag of machines kunnen denken. In plaats van een analytische betekenisanalyse van de dubbelzinnige begrippen ‘machine’ en ‘denken’ kiest hij ervoor de vraag te operationaliseren. Hij definieert hiertoe een imitatiespel welke er in essentie op neerkomt dat een menselijke ondervrager moet achterhalen of zijn of haar gesprekspartner een mens of een machine is. De gesprekspartner bevindt zich in een andere kamer dan de ondervrager waardoor visueel contact voorkomen wordt. De vragen en antwoorden worden bijvoorbeeld via een telex uitgewisseld zodat de ondervrager de stem van de gesprekspartner ook niet kan horen. Turing operationaliseert ook het concept ‘machine’ door deze te vervangen door ‘universeel programmeerbare digitale computer’. Op deze manier sluit hij experimentele methoden, biologische methoden en voortplanting uit zonder daarmee aan een bepaalde ingenieurspraktijk voor machinebouw de voorkeur te geven.

Het imitatiespel trekt een scherpe lijn tussen fysieke en intellectuele vermogens op een manier die juist de fysieke vermogens irrelevant maakt. Turing ziet dit als een groot voordeel van het imitatiespel omdat het voor de vraag naar intelligentie immers niet van belang is of een computer er al dan niet menselijk uitziet. Bovendien willen we in dit geval een computer niet afrekenen op het feit dat deze bijvoorbeeld niet kan schitteren in schoonheidswedstrijden.

Turing formuleert binnen zijn nieuwe operationele context een eenduidige en falsificeerbare claim als reactie op de volgens hem betekenisloze originele vraag: ‘Binnen 50 jaar zal het mogelijk zijn om computers met een bepaalde geheugencapaciteit te programmeren welke het imitatiespel zo goed spelen dat de gemiddelde ondervrager niet meer dan 70 procent kans heeft om na 5 minuten de juiste identificatie te maken’.

Het is volgens Turing onmogelijk om de oorspronkelijke vraagstelling geheel te vermijden omdat niet iedereen zijn operationalisering zal accepteren. Dit blijkt inderdaad uit zijn behandeling van een negental bezwaren tegen zijn opvattingen over kunstmatige ofwel artificiële intelligentie. Ieder van deze bezwaren zal ik hieronder toelichten.

*Het theologische bezwaar* - Denken is een functie van de onsterfelijke ziel van de mens. God verleent dergelijke zielen alléén aan mensen en dus niet aan machines. Machines kunnen daarom niet denken. Hoewel Turing weinig onder de indruk is van theologische bezwaren geeft hij toch een reactie in theologische termen. Het bezwaar lijkt geen recht te doen aan de veronderstelde almacht van God. Waarom zou God immers geen ziel kunnen verlenen aan bepaalde diersoorten of zelfs aan specifieke machines wanneer hij dat gezien de omstandigheden gepast zou vinden? De mens als machinebouwer is in het tweede geval niets meer dan een instrument van Zijn wil.

*Het ‘kop in het zand’ bezwaar* – De gevolgen van denkende machines zijn zo verschrikkelijk

dat we beter kunnen menen en geloven dat machines niet kunnen denken. Turing geeft aan dat de gedachte dat de mens noodzakelijk superieur is aan dit bezwaar ten grondslag ligt. Het bezwaar houdt verder verband met het eerder besproken theologische argument en is vooral populair bij intellectuelen die het denkvermogen hoger waarderen dan andere vermogens. Turing vindt het bezwaar niet substantieel genoeg om er verder inhoudelijk op in te gaan.

*Het wiskundige bezwaar* – Volgens de stelling van Gödel bevat ieder consistent en voldoende krachtig[1] logisch systeem uitspraken die binnen het systeem noch bewezen noch weerlegd kunnen worden. Er bestaan verschillende soortgelijke stellingen zoals ondermeer 'de stelling van Turing'. Deze stelling heeft direct betrekking op machines in plaats van op logische systemen en is daarom meer geschikt voor het beschrijven van het wiskundige bezwaar. In feite heeft de stelling van Turing alléén betrekking op een bijzonder soort van machines welke hierna steeds aangeduid zullen worden als 'Turing machines'. Turing zelf omschrijft dit type machines als 'digitale computers met oneindige geheugencapaciteit'. De stelling van Turing luidt nu dat er voor iedere Turing machine minimaal één 'ja/nee' vraag bestaat waarop de machine ofwel géén ofwel een fout antwoord geeft[2]. Het soort 'ja/nee' vragen waaraan hier gedacht moet worden zijn vragen waarbij een bepaalde Turing machine wordt gevraagd om het 'ja/nee' antwoordgedrag van een andere nauwkeurig gespecificeerde Turing machine te voorspellen. Het wiskundige bezwaar komt erop neer dat de stelling van Turing laat zien dat machines[3] onderhevig zijn aan beperkingen waaraan het menselijke intellect zelf niet onderhevig is. Machines kunnen dus niet denken zoals mensen dat kunnen. Turing merkt als reactie op dit bezwaar op dat helemaal niet bewezen is dat de genoemde beperkingen voor machines niet gelden voor de menselijke geest. Ieder mens geeft immers ook regelmatig onjuiste antwoorden op bepaalde 'ja/nee' vragen.

*Het bezwaar op basis van bewustzijn* – Geen enkel mechanisme dat louter symbolen manipuleert lijkt gevoelens als plezier, teleurstelling of woede te kunnen bezitten. Het imitatiespel van Turing let uitsluitend op extern waarneembaar gedrag en gaat daarmee voorbij aan de eigenlijke vraag of machines (zelf)bewustzijn ofwel subjectieve ervaringen kunnen hebben. De enige manier om erachter te komen of machines kunnen denken zou zijn om zelf die machine te zijn en vervolgens te voelen dat men daadwerkelijk denkt. Deze positie leidt volgens Turing tot een problematische vorm van solipsisme omdat het niet langer mogelijk is om te bewijzen dat een andere entiteit dan jezelf beschikt over (zelf)bewustzijn. Turing stelt verder dat de aanhangers van dit bezwaar toch zullen moeten toegeven dat er daadwerkelijk sprake is van 'echt begrip' in plaats van slechts 'vernuftige machinerie' wanneer zou blijken dat een machine zeer bevredigende, subtiele en verfijnde antwoorden kan geven op bijvoorbeeld geraffineerde gevoelsvragen over poëzie.

*Het bezwaar op basis van verschillende onvermogens* – Dit bezwaar stelt dat er vermogens

---

[1] Een logisch systeem is in dit verband 'voldoende krachtig' wanneer zij de rekenkunde, of beter gezegd de Peano arithmetiek voor de natuurlijke getallen (PA), omvat. Dit blijkt o.a. uit het door Gödel gegeven bewijs.

[2] De stelling waarop Turing hier doelt, staat in de literatuur bekent als het zogenaamde 'halting' probleem.

[3] De hier gehanteerde overgang van 'Turing machines' naar 'machines' kan gerechtvaardigd worden op basis van de Church-Turing these. Deze these stelt dat de klasse van intuïtief berekenbare functies gelijk is aan de klasse van Turing machine berekenbare functies. Ieder probleem dat door een willekeurige machine kan worden opgelost zou dus ook door een Turing machine kunnen worden opgelost ofwel ieder voorstelbaar algoritme zou implementeerbaar zijn op een Turing machine. De laatste jaren staat deze these echter steeds meer onder druk.

zijn die een machine nooit zal kunnen bezitten zoals initiatief tonen, een gevoel voor humor hebben, goed van kwaad onderscheiden, vergissingen maken, verliefd worden, genieten van aardbeienijs, zichzelf als onderwerp van gedachten nemen, een even grote verscheidenheid in gedrag vertonen als de mens of iets volkomen nieuws doen. Volgens Turing zijn deze beweringen gebaseerd op inductie. Mensen extrapoleren hun ervaringen met machines uit het verleden naar de toekomst. Turing vraagt zich af hoe sterk een beroep op inductie in dit verband kan zijn. Niet ieder onderwerp leent zich voor inductie. Bovendien moet een zeer groot deel van een domein onderzocht worden om tot echt betrouwbare resultaten te komen. Hij stelt verder dat de beperkte opslagcapaciteit van de huidige machines verantwoordelijk is voor veel van de genoemde onvermogens. Turing gaat op enkele van deze onvermogens nader in. Machines kunnen volgens hem juist wel (functionele) fouten maken, zoals bijvoorbeeld een machine die inductie toepast. Ook is het voorstelbaar dat een machine zichzelf als onderwerp neemt. Tenslotte beweert Turing dat machines net zo divers gedrag als de mens kunnen gaan vertonen zodra zij de beschikking krijgen over veel meer opslagruimte.

*Het bezwaar van Lady Lovelace* – Een machine zou ons nooit kunnen verrassen door iets daadwerkelijk nieuws te doen omdat een machine alleen maar opdrachten uitvoert die de mens haar aanreikt. Turing vindt dit argument niet erg sterk omdat (zoals aangegeven door Hartree) niet uitgesloten is dat wij een zelfdenkende of zelflerende machine kunnen maken. Turing stelt verder dat wij als mensen er ook niet zeker van zijn dat onze ‘originele vondsten’ niet het gevolg zijn van het strikt toepassen van algemene procedurele regels. Bovendien zegt Turing dat hij regelmatig door machines wordt verrast. De gedachte dat machines ons niet zouden kunnen verrassen is volgens hem gebaseerd op de onjuiste aanname dat alle (logische) gevolgen van een bepaald feit voor ons onmiddellijk bekend zijn zodra het feit zelf bekend is.

*Het argument op basis van de continuïteit in het zenuwstelsel* – Het zenuwstelsel is een continu systeem welke niet nagebootst kan worden door een discrete toestandsmachine. Volgens Turing is dit geen doorslaggevend argument omdat in het imitatiespel een discrete machine juist wel een continu systeem adequaat kan nabootsen. Zo kan een digitale computer (discreet) tijdens een Turing test bijvoorbeeld een zogenaamd differentiaal analyse instrument (continu) voldoende nauwkeurig simuleren om het de ondervrager heel erg lastig te maken.

*Het argument op basis van het informele karakter van gedrag* – Het is onmogelijk om een vaste verzameling van gedragsregels te definiëren waarmee iemand in iedere denkbare situatie kan bepalen hoe te handelen. Machines reageren op basis van dergelijke vaste regels en daarom kan de mens dus geen machine zijn. Turing stelt voor om te denken in termen van natuurwetten die het gedrag van iemand reguleren (gedragswetten) in plaats van expliciete gedragsregels waarmee iemand zijn leven bewust reguleert. Wanneer we dit doen lijkt het genoemde onderscheid tussen mensen en machines te verdwijnen omdat beide begrepen kunnen worden als entiteiten gereguleerd door gedragswetten. Bovendien kunnen we ons niet zo gemakkelijk overtuigen van het niet bestaan van gedragswetten als van het niet bestaan van gedragsregels.

*Het argument gebaseerd op buitenzintuiglijke waarneming* – Volgens Turing bestaat er overtuigend statistisch bewijs voor fenomenen als telepathie. Wanneer we de mogelijkheid

van buitenzintuiglijke waarneming accepteren moeten we er eveneens vanuit gaan dat juist bij het denken fenomenen als telepathie een rol spelen. Door nu vragen te stellen die helderziendheid bij de gesprekspartner vereisen kan de ondervrager tijdens het imitatiespel gemakkelijker machines 'ontmaskeren'. Machines kunnen immers niet meer dan willekeurig raden. Een ondervrager zou door gebruik te maken van telepathie zelfs machines als gesprekspartner kunnen ontmaskeren zonder enige vraag te stellen. Wanneer telepathie is toegelaten wordt het dus van belang om bijvoorbeeld tijdens de test gebruik te maken van 'telepathie vrije kamers'.

**Discussievragen**

1. Turing noemt in zijn artikel het imitatiespel ook wel een 'criterium' of 'test'. Is het kunnen doorstaan van de Turing test een noodzakelijke voorwaarde voor het mogen claimen van een dispositie tot intelligent gedrag? Is het kunnen doorstaan van de Turing test (ook) een voldoende voorwaarde voor het mogen claimen van een dispositie tot intelligent gedrag?

2. Het is nog maar de vraag of het kunnen vertonen van intelligent gedrag altijd samengaat met het bezitten van (zelf)bewustzijn. Niet voor niets moet er steeds een onderscheid gemaakt worden tussen enerzijds het 'makkelijke' probleem (nabootsen van intelligent gedrag) en anderzijds het 'moeilijke' probleem (begrijpen van de aard van het bewustzijn ofwel het qualia probleem). Daarom dezelfde vragen als onder [1] voor het mogen claimen van het hebben van (zelf)bewustzijn.

3. De kritiek van Turing op het mathematische bezwaar lijkt overtuigend. Er bestaat echter een iets scherpere formulering van het mathematische bezwaar welke ik hieronder kort zal weergeven. Is deze scherpere formulering net zo eenvoudig weerlegbaar?

**Stelling:** De mens is geen machine. **Bewijs (uit het ongerijmde):** Stel dat de mens wél een machine is zodat iedere persoon samenvalt met een bepaalde Turing machine. Laat P een menselijke persoon zijn welke samenvalt met Turing machine M. M is een Turing machine en daarom onderhevig aan 'de wet van Turing'. Er bestaat dus een algoritmische procedure om vragen te construeren welke M ofwel fout ofwel niet zal beantwoorden en waarvoor bovendien geldt dat de mens het juiste antwoord wel kan achterhalen. Stel nu dat P (of iemand anders) dit algoritme op M toepast en als gevolg hiervan de vraag V terugkrijgt. P kan nu (eventueel met hulp van anderen) het juiste antwoord op V achterhalen terwijl M de vraag V ofwel fout ofwel niet beantwoordt. P is dus niet gelijk aan M. We hadden echter aangenomen dat P met M samenvalt zodat een tegenspraak verkregen is. De oorspronkelijke aanname moet dan ook worden verworpen. De mens is dus geen machine. QED

4. Er zou beweerd kunnen worden dat de Turing test helemaal geen intelligent gedrag meet. Zij zou 'slechts' meten of een gesprekspartner al dan niet goed is in het doorstaan van Turing tests. Vergelijk in dit verband IQ testen waarvan ook vaak beweerd wordt dat zij geen intelligentie meten maar alleen iemands vaardigheid in het succesvol afleggen van IQ tests. Ook Turing zelf sluit in zijn artikel niet uit dat er strategieën bestaan om het imitatiespel te 'winnen' zonder daadwerkelijk intelligent gedrag te hoeven vertonen. Is deze kritiek terecht?

5. Is het al dan niet hebben van een lichaam irrelevant voor het al dan niet kunnen vertonen van intelligent gedrag? Vergelijk in dit verband de ideeën van Merleau-Ponty over ondermeer het 'corps sujet'. In de wereld van het onderzoek naar artificiële intelligentie (AI) is de laatste jaren ook een trend waarneembaar van steeds minder onderzoek naar intelligente logische

systemen (zonder lichaam) naar steeds meer onderzoek naar intelligente robots (met lichaam). De interesse van AI onderzoekers voor virtuele werelden (waarbij elk karakter een lichaam heeft en zich in een ruimtelijke wereld voortbeweegt) past ook in deze ontwikkeling.