

# **Computing machinery and Intelligence**

## **A. M. Turing**

*Samengevat door: Matthijs Melissen*

Ik stel voor om de vraag “Kunnen machines denken?” te behandelen door te kijken naar een zogenaamd “imitatiespel”. Hiervoor nemen we een persoon en een computer. Nu laten we een tweede persoon in een andere ruimte via op afstand bestuurbare printers vragen stellen aan zowel de computer als de persoon. Zowel de mens als computer proberen te doen alsof ze een mens zijn. Als de ondervrager net zo slecht onderscheid kan maken tussen de mens en de computer, als dat hij dat zou kunnen tussen een ondervraagde man en vrouw, kunnen we stellen dat de machine kan denken.

De voorgestelde methode heeft het voordeel dat we hiermee alleen naar de fysieke en niet naar de intellectuele capaciteiten van een mens en een machine kijken. Het spel kan wel bekritiseerd worden omdat de computer de mens moet proberen na te doen en niet andersom. Een mens zal een computer ook niet makkelijk kunnen nabootsen. Misschien denkt een computer ook wel, maar op een totaal andere manier dan mensen doen. Het spel is dus wel geschikt om de intelligentie van computers vast te stellen, maar niet om te kunnen ontkennen dat een computer intelligentie bezit. Tot slot kunnen we ons afvragen of de machine iets anders moet doen dan een mens imiteren om zijn intelligentie te bewijzen, maar dat lijkt onwaarschijnlijk.

We staan als machines alleen de elektronische digitale computer toe, want bijvoorbeeld gekloonde mensen willen we niet onder de definitie van machine laten vallen. Dit lijkt een grote beperking, maar dat is het niet, want nog steeds kunnen we de intelligentie van computers met dit spel bewijzen. Pas als digitale computers niet in staat zijn dit spel te winnen, wat ik niet geloof, wordt het tijd om naar andere typen computers te gaan zoeken die dit misschien wel kunnen.

We zijn niet geïnteresseerd in de vraag of alle computers in de wereld het spel goed kunnen spelen, of in de vraag of er hedendaagse computers zijn die dit kunnen, maar of er denkbeeldige computers zijn die het kunnen.

Een digitale computer kan in principe alle bewerkingen uitvoeren die ook door een menselijke “computer” uitgevoerd kunnen worden. Zowel een menselijke als een digitale computer worden geacht een serie vaste regels uit te voeren bij hun berekeningen. Een digitale computer bestaat over het algemeen uit drie delen: opslagruimte, een uitvoereenheid en een controle-eenheid. De opslagruimte komt overeen met het geheugen of kladpapier van de menselijke computer. De uitvoereenheid is het onderdeel dat de individuele bewerkingen waar een berekening uit bestaat uitvoert. De controle-eenheid zorgt ervoor dat de regels correct en in de goede volgorde worden uitgevoerd. Die instructies bestaan uit een serie getallen, waaruit de te gebruiken getallen en de bewerking

die op de getallen die moet worden uitgevoerd volgen. Een instructie kan ook de opdracht bevatten op een andere positie in de serie regels verder te gaan. Het ontwerpen van instructietabellen wordt “programmeren” genoemd. Een interessante variant op de digitale computer is een digitale computer met een toevalselement. De functie van zo’n component is te vergelijken met het gooien van een dobbelsteen. Van zo’n machine wordt wel eens gezegd dat deze vrije wil heeft. Het is normaal gesproken niet mogelijk te achterhalen of een machine zo’n toevalselement heeft, want elk resultaat dat met zo’n toevalselement bereikt kan worden, is ook op andere manieren te bereiken. We kunnen ons een “oneindige-capaciteit-computer” voorstellen. Hieraan voegen we steeds minimaal net zo veel opslagruimte toe als nodig, zodat deze virtueel een oneindige opslagruimte heeft.

Charles Babbage ontwierp zo’n machine, genaamd de “Analytical Engine”, maar hij bouwde deze nooit af. De machine zou vanwege zijn lage snelheid overigens ook niet erg aantrekkelijk zijn geweest. Dit was een puur mechanisch apparaat, dus blijkbaar is het geen vereiste dat een digitale computer elektrisch werkt.

Een digitale computer zoals in de vorige paragraaf is een voorbeeld van een “discrete toestand-machine”. Dit is een machine waarvan de toestanden dermate verschillend zijn dat ze niet met elkaar verward kunnen worden. De toestand op elk moment wordt geheel bepaald door de invoer en vorige toestand.

Computers hebben bijzonder veel toestanden. De logaritme met grondtal twee van het aantal toestanden wordt de “opslagcapaciteit” van de machine genoemd. Als twee machines worden samengevoegd moeten de capaciteiten hiervan worden opgeteld om de capaciteit van de nieuwe server te krijgen.

Als we de instructies en invoer van een discrete toestand-machine kennen, kunnen we voorspellen wat deze zal doen. We kunnen elk programma dat we op een discrete toestand-machine uitvoeren dus ook op een digitale computer uitvoeren, als deze maar over voldoende snelheid en opslagcapaciteit en een geschikt programma beschikt. We zeggen dat de digitale computer een universele machine is. Bovendien kunnen we concluderen dat alle computers equivalent zijn. We kunnen de vraag “kunnen machines denken?” dus vervangen door de vraag “kunnen discrete toestand-machines denken”. Dit kunnen we met het imitatiespel testen.

Ik geloof dat in 50 jaar tijd een computer het imitatiespel zo goed kan spelen dat een gemiddelde ondervrager minder dan 70% kans heeft om de computer te ontmaskeren. Dat is slechts mijn hypothese, geen bewezen stelling. Ik zal nu eerst enkele argumenten tegen deze stelling bespreken.

- De theologische tegenwerping

Deze tegenwerping luidt als volgt: Denken is een functie van de onsterfelijke menselijke ziel. God heeft een onsterfelijke ziel aan elk mens gegeven, maar niet aan dieren of mensen. Daarom kan geen enkel dier of machine kan denken.

Naar mijn mening is er een groter verschil tussen levend wezen en niet-levend wezen dan tussen mens en dier, dus dieren zouden beter bij de mens ingedeeld kunnen worden.

Waarom zou God bovendien niet de vrijheid hebben om een computer een ziel te verlenen als Hij ziet dat dat mooi uitkomt? Waarschijnlijk zou hij die macht alleen willen gebruiken als de computer voldoende ontwikkeld zou zijn om met het bezit van een ziel

om te gaan, maar dat is geen onvervulbare vereiste. Over die omstandigheden gaan we het volgend hoofdstuk hebben. Tot slot, als we zulke machines proberen te maken, nemen we niet respectloos Zijn macht om zielen te maken over, want we zijn instrumenten van Zijn wil.

- De hoofd-in-het-zand tegenwerping

Dit argument komt neer op de uitspraak “Denkende machines hebben vervelende gevolgen. Laten we daarom hopen dat ze onmogelijk zijn.” Dit argument is onvoldoende substantieel om het ongelijk te hoeven bewijzen.

- De wiskundige tegenwerping

Er bestaat wiskundige logica die kan worden gebruikt om de beperkingen van discrete toestand-machines aan te tonen. We weten echter niet of mensen niet ook die beperkingen hebben. Degenen die dit argument aanvoeren zullen bereid zijn het imitatiespel aan te nemen als een basis voor discussie. Degenen die geloven in de twee tegenwerpingen hiervoor zijn waarschijnlijk niet geïnteresseerd in welk criterium dan ook.

- De bewustzijn-tegenwerping

Volgens sommigen is bewustzijn een vereiste voor intelligentie. Dat is echter problematisch omdat je dan alleen van je zelf kunt zeggen dat je denkt en niet van anderen. De meeste aanhangers van deze tegenwerping zullen liever hun standpunt willen verlaten dan in deze solipsistische positie terecht te komen. Dan zullen ze waarschijnlijk onze test willen accepteren.

- Argumenten met betrekking tot diverse onmogelijkheden

Deze argumenten hebben de vorm “Ik geef toe dat je machine alles kan wat je hebt genoemd, maar het zal nooit X kunnen doen”. Deze uitspraken zijn vooral gebaseerd op het feit dat men nog nooit een machine tegengekomen is die dit kan, en niet op een principiële onmogelijkheid. We kunnen bijvoorbeeld ook in computers inbouwen dat ze fouten maken. Een machine kan verder ook over zichzelf nadenken, bijvoorbeeld door een programma aan te passen aan de hand van de resultaten van zijn eigen gedrag. Het is ook niet waar dat een machine weinig divers gedrag kan vertonen; dat komt vooral door de beperkte opslagcapaciteit.

- Lady Lovelace’s tegenwerping

Lady Lovelace (1842) schreef over de Analytical Engine van Babbage dat het niets creëert, maar alleen uitvoert wat wij geordend aanbieden aan deze machine. Dit betekent echter niet dat het niet mogelijk is een elektrisch apparaat te ontwerpen wat in staat is zelf na te denken. Een variant op deze tegenwerping is dat een computer nooit “iets echt nieuws kan doen” of “kan verassen”. Er is echter niets waarvan we zeker weten dat het nieuw is, want wie zegt dat menselijk “nieuw” gedrag niet voortkomt uit aangeleerd gedrag. Machines kunnen de programmeur bovendien wel degelijk verassen, omdat de programmeur vaak niet voldoende rekenkracht heeft om de uitkomst van zijn programma te voorspellen.

- Het argument van een continu zenuwstelsel

Iemand zou kunnen stellen dat een discrete toestand-machine niet het gedrag van het

zenuwstelsel kan imiteren, omdat het zenuwstelsel niet discreet maar continu werkt. De ondervrager in het imitatiespel kan echter geen onderscheid maken tussen een discrete en de continue machine, omdat ze hetzelfde type antwoorden kunnen geven.

- Het argument van informeel gedrag

Dit argument gaat als volgt: als elke persoon een begrensde verzameling regels had waarmee hij zijn gedrag zou bepalen, zou hij niet beter zijn dan een machine. Zulke regels bestaan niet, dus mensen zijn beter dan machines. Dat is echter geen geldige logische redenering. We kunnen bovendien niet aantonen dat er geen gedragsregels voor mensen zijn.

- Het argument van buitenzintuiglijke waarneming

Buitenzintuiglijke waarneming, zoals telepathie, kan een storende werking op het imitatiespel hebben. Dit probleem kunnen we oplossen door te zeggen dat de deelnemers zich in een “telepathie-vrije ruimte” moeten bevinden.

Ik heb geen overtuigende positieve argumenten voor mijn standpunten, hooguit uitspraken die mijn standpunten waarschijnlijker maken.

Misschien kunnen we “superkritische” machines maken, zodat ze na een startsignaal een steeds uitbreidende stroom gebeurtenissen opwekken.

Als we de hersenen mechanisch ontleden zoals we een ui pellen, houden we dan uiteindelijk de kern van de hersenen over, of uiteindelijk een lege schil? In het laatste geval is het brein geheel mechanisch.

De enige echt overtuigende ondersteuning van mijn uitspraak zou een computer zijn die aan het eind van de eeuw in staat is het imitatiespel succesvol voor de computer te laten spelen. Dat is hoofdzakelijk een programmeerprobleem, dat opgelost kan worden als er vorderingen in het ontwerp van computers gemaakt worden.

Om de hersenen na te kunnen bootsen moeten we bekijken hoe de hersenen geworden zijn tot wat ze zijn. We onderscheiden drie componenten: de toestand bij de geboorte, het onderwijs wat het gekregen heeft en andere ervaringen waar het aan onderworpen is. In plaats van volwassen hersenen kunnen we ook kinderhersenen simuleren en deze onderwijzen. Door de ontwerper te laten kijken welk ‘kindermodel’ het beste kan leren krijgen we een soort van evolutie, alleen sneller en doelmatiger. Straf en beloning kunnen onderdeel uitmaken van het leerproces, maar dit is te weinig om het leerproces voldoende te sturen. De lerende machine moet regels bevatten die niet kunnen veranderen, maar ook regels die de machine zelf kan veranderen.

De meningen verschillen over hoe complex de machine die nog niet onderwezen is moet zijn. Een simpele computer is makkelijker te bouwen, een ingewikkelde computer makkelijker te onderwijzen.

Een belangrijke eigenschap van de machine is dat de leraar vaak niet weet wat er zich in de machine afspeelt, maar tot op zekere hoogte wel het gedrag van de “leerling” kan voorspellen. Op deze manier wordt ook de menselijke eigenschap van het kunnen maken van fouten in de computer geïntroduceerd, omdat iets dat geleerd is geen 100% zeker resultaat geeft.

Waarschijnlijk is het verstandig om een toevalselement in een lerende machine te stoppen, omdat dat erg handig is bij het zoeken naar een oplossing voor een probleem.

We hopen dat computers uiteindelijk in alle intellectuele gebieden net zo goed of beter als mensen worden, maar welk gebied is het beste om mee te beginnen? Veel mensen denken aan een abstracte activiteit als schaken, anderen aan het begrijpen en spreken van Engels. Ik weet niet wat het goede antwoord is, maar ik denk dat beide benaderingen uitgeprobeerd moeten worden.

We kunnen maar kort vooruitzien, maar we kunnen een hoop zien dat gedaan moet worden.

Turing, Alan. 1950. Computing Machinery and Intelligence. *Mind, A Quarterly Review of Psychology and Philosophy*. vol 59, 433 – 460.