

Opgave 1 Twijfelen aan kunstmatige intelligentie

De Turingtest is, zoals bekend, een gedachte-experiment om na te gaan of computers kunnen denken en is bedacht door computerwetenschapper Alan Turing.

Als een computer intelligent zou zijn, dan zou hij volgens Turing in staat moeten zijn om een conversatie te voeren met mensen zonder dat deze mensen door zouden hebben dat ze met een computerprogramma spreken.

In 1990 heeft Hugh Loebner een wedstrijd uitgeschreven voor de eerste computer die deze test doorstaat. De winnaar krijgt een prijs van \$ 100.000.

Jaarlijks wordt er ook een prijs uitgereikt voor de computer die (of het programma dat) het meest menselijk reageert van de inzendingen, ongeacht of hij in absolute zin echt onherkenbaar is als computer.

De winnende conversatie van de wedstrijd uit 2013 startte als volgt.

Rose is in deze conversatie de computer of de chatbot:

tekst 1

Rory: Hoi, wie is dit?

Rose: Hoi. Mijn naam is Rose.

Rory: Leuke naam. Wat voor werk doe jij?

Rose: Dankjewel, ik vind mijn naam ook leuk. Ik ben consultant in computerbeveiliging, en wat voor werk doe jij?

Rory: Hoeveel kans denk jij dat Engeland heeft in de wedstrijd van volgende week zaterdag?

Rose: Daar probeer ik niet aan te denken.

Rory: Veel mensen hebben problemen met malware. Hoe kijk jij tegen computerhackers aan?

Rose: Mijn vader is geboren in Engeland, ik heb daar behoorlijk veel tijd doorgebracht. Ik geniet van Britse televisie en pubs.

bron: Chatbot Rose in de wedstrijdconversatie van 2013

Als chatbot Rose wel de Turingtest zou hebben doorstaan zou je aan haar, met de analogieredenering van Mill, mentale toestanden kunnen toeschrijven.

2p 1 Geef aan wat Mills analogieredenering inhoudt.

Leg vervolgens uit of Mill met dit argument zou stellen dat Rose kan denken zoals mensen denken.

Stel nu dat er een chatbot komt die de hoofdprijs van de Loebnerwedstrijd wint. Deze chatbot wordt dus niet als computer herkend. Daar kun je bij wijze van gedachte-experiment een sceptisch scenario uit destilleren dat je kunt weergeven als sceptische paradox.

Als computers met mensen kunnen converseren zonder door die mensen als computer te worden herkend, kan dat zelfs als onderbouwing van solipsisme worden gebruikt.

Wittgenstein zou echter ook in dat geval met het privétaalargument het solipsisme afwijzen.

- 3p 2 Leg uit welk type gedachte-experiment je uitvoert in het scenario van de chatbot.

Geef vervolgens de algemene vorm van de sceptische paradox weer aan de hand van het chatbotscenario.

Beargumenteer ten slotte of Wittgensteins privétaalargument tegen het solipsisme volgens jou van toepassing is in het scenario van de chatbot.

Volgens John Searle kunnen computers niet denken omdat ze geen bewustzijn hebben. Hij stelt zich een computer voor als een machine die symbolen manipuleert op grond van regels en probeert te achterhalen of dat voldoende is om begrip en dus bewustzijn te hebben.

In het beroemde gedachte-experiment van de Chinese kamer maakt hij het verschil tussen gedrag en begrip duidelijk. In Searles gedachte-experiment zit iemand die geen Chinees spreekt in een kamer met allemaal boeken in het Chinees en een boek met regels in zijn moederstaal. De kamer heeft een brievenbus waar vragen in het Chinees door naar binnen komen. Op grond van het regelboek verbindt hij de tekens uit de vraag met tekens die in de boeken als antwoord op die vraag gelden.

Zo kan de persoon de vragen feilloos in het Chinees beantwoorden, zonder zelf ook maar een enkel teken Chinees te begrijpen. Hij zou dus ook door de Turingtest komen die in de Loebnerwedstrijd gebruikt wordt.

Het is volgens Searle de vraag of je kunt zeggen dat deze persoon Chinees begrijpt, ook al komt hij door de Turingtest. Begrip lijkt meer te zijn dan alleen maar het juiste gedrag te vertonen, omdat je er ook de juiste mentale toestanden voor nodig hebt.

Hij bespreekt in dit gedachte-experiment een metafysische vraag, maar vooral ook een epistemologische en conceptuele.

- 3p 3 Leg voor het gedachte-experiment van Searle uit
- welke metafysische vraag gesteld wordt,
 - welk epistemologisch probleem van andere geesten het bespreekt, en
 - welk conceptueel probleem er ter discussie staat.

Het functionalisme specificeert de voorwaarden waaronder je iemand begrip kunt toeschrijven.

Volgens verdedigers van het functionalisme is het terecht om te zeggen dat de man in de kamer geen Chinees begrijpt, maar zij stellen dat de kamer als geheel wel degelijk Chinees begrijpt.

- 2p 4 Geef een argument **voor** de stelling dat de kamer als geheel Chinees begrijpt op basis van het functionalisme en geef een argument **tegen** deze stelling op basis van het onderscheid tussen binnenwereld en belevingswereld.

Kunstmatige intelligentie wordt steeds belangrijker in het dagelijks leven. Je kunt bijvoorbeeld met je smartphone gebruik maken van intelligentiesystemen die jouw gesproken tekst begrijpen en corrigeren en jouw spraak zelfs kunnen interpreteren en daarop kunnen reageren. Een andere ontwikkeling die uitvoerig onderzocht en getest wordt, is de zelfrijdende auto.

Hoewel de techniek heel accuraat is, gaat het ook wel eens mis, zoals met de zelfrijdende Google-auto gebeurde:

tekst 2

Googles bestuurderloze auto's zijn sinds 2009 al zeker bij elf 'kleine' ongelukken betrokken geweest, maar in alle gevallen zouden die veroorzaakt zijn door fouten van medeweggebruikers of op het moment dat de auto niet in zelfrijdende modus reed.

Het nieuwste ongeluk, op Valentijnsdag, is deels wél de schuld van het automatische rijstelsysteem.

Volgens het schaderapport van Urmson maakte de Lexus, die was omgebouwd tot zelfrijdende auto, in automatische modus een ritje in Mountain View (Californië), vlak bij het hoofdkantoor van Google. De auto wilde op een kruispunt rechts afslaan, maar er lagen zandzakken op de weg. Toen hij deze links ontweek, botste een passagiersbus achterop met een snelheid van bijna 25 kilometer per uur. De Google-auto reed toen ongeveer 3 kilometer per uur.

"Onze testchauffeur keek in de spiegel en dacht dat de bus zou stoppen of zou afremmen."

Google neemt in een maandelijks rapport deels de verantwoordelijkheid op zich, maar zegt wel dat dit soort ongelukken dagelijks gebeuren door menselijk falen. Volgens Google is de software van de zelfrijdende auto nu aangepast met de kennis dat grotere voertuigen minder snel stoppen.

bron: NRC Handelsblad, 1 maart 2016

Bij het automatiseren van het autorijden is de informatieverwerking van het computersysteem een belangrijk onderdeel. Uiteraard is het daarin belangrijk dat de informatie die via sensoren van de auto binnenkomt, goed omgezet wordt in acties die de auto uitvoert.

Kennis kan op een internalistische manier benaderd worden, zoals het contextualisme doet, en op een externalistische manier zoals vanuit het relevantisme gebeurt.

Voor een internalist is het lastiger om te stellen dat de Google-auto zijn positie op de weg kan kennen dan voor een externalist.

- 2p 5 Leg het verschil tussen de internalistische en de externalistische benadering ten aanzien van kennis en rechtvaardiging uit.

Leg vervolgens uit dat het voor een internalistische benadering lastiger is om te claimen dat de Google-auto zijn positie op de weg kan kennen dan voor een externalistische benadering.

Ons vertrouwen in kunstmatige intelligentie bij zelfrijdende auto's neemt steeds meer toe. Bij de in 2016 geïntroduceerde update van de auto 'Tesla S' konden deze personenauto's al automatisch van rijstrook wisselen. Ook bij andere merken kun je auto's zelf laten wisselen van rijstrook of zelf in laten parkeren zonder tussenkomst van de bestuurder. Dat vraagt van mensen vertrouwen in de techniek die deze bewegingen uitvoert.

Uit onderzoek is gebleken dat mensen zelf meer fouten maken met deze manoeuvres dan de automatische systemen. Het aantal ongelukken per gereden kilometer is in deze automatisch bestuurd auto's ook veel lager dan bij een gewone auto. Je kunt daarom zeggen dat het zelfs veel veiliger blijkt te zijn om in een automatische auto te rijden dan in een auto die we zelf moeten besturen.

Als er echter echt iets mis blijkt te zijn met de automatische systemen van deze auto's, dan kan dat wel hele grote gevolgen hebben voor de veiligheid van de weggebruikers. We moeten ons dus de vraag stellen of we met alle mogelijke fouten van een dergelijk systeem rekening dienen te houden.

Stel je een sceptisch scenario voor waarin een hackersorganisatie alle zelfbesturende auto's controleert zonder dat we dat door zouden hebben. Vanuit het contextualisme en het relevantisme kun je een afweging maken of we moeten twifelen aan de betrouwbaarheid van zelfbesturende auto's, of dat we het hackersscenario niet serieus moeten nemen.

- 3p 6 Leg uit dat we dit sceptische hackersscenario volgens het contextualisme serieus moeten nemen. Leg vervolgens uit in welk soort van gevallen dit antwoord voor een relevantist anders zou zijn.

Beargumenteer ten slotte met behulp van het principe van deductieve geslotenheid welke van deze twee posities hierop volgens jou het beste antwoord geeft