

ingen forskjell i det hele tatt, og eksperter kunne ikke gjenkjenne opptil en tredjedel av tekstene skrevet av AI som kunstig genererte (Gao et al., 2022). Dette kan få etiske og (vitenskaps-) politiske konsekvenser, en dimensjon vi så vidt begynner å forstå. (Mer om dette i kapittel 13.)

Det kinesiske rommet: en innvending mot den funksjonalistiske tilnærmingen

Turingtesten bygger på antakelsen om at hvis den korrekte syntaktiske formen er oppfylt, vil den semantiske betydningen komme automatisk. En velkjent innvending mot akkurat den funksjonalistiske tilnærmingen og ideen om at datamaskiner kan tenke, er tankeeksperimentet «**Det kinesiske rommet**». John Searle (1932-) argumenterer mot det han kaller «sterk kunstig intelligens». Den skiller seg fra «svak kunstig intelligens». Svak kunstig intelligens betyr at datamaskiner i beste fall blir sett på som nyttige verktøy som hjelper til med å løse en kompleks oppgave (Searle, 1980, s. 417). Et godt eksempel på svak kunstig intelligens kan være ekspertsystemer som programvare for å oppdage kreftsvulster. Sterk kunstig intelligens er imidlertid noe helt annet. Ifølge Searle betyr denne typen kunstig intelligens at datamaskiner eller andre kunstige systemer (som bruker en datamaskin) er intelligente i den forstand at de faktisk er i stand til å forstå noe slik vi mennesker gjør. De er et sinn i ordets sanneste betydning.

I tankeeksperimentet sammenligner Searle informasjonsberegningen til en datamaskin med den semantiske språkforståelsen til mennesker (Searle, 1980). Eksempelet er satt opp slik at det vil bestå

Turingtesten, men det viser samtidig hva som er problematisk med den.

Du kan forestille deg det slik: Du sitter i et rom og behandler kinesiske tegn ved å følge instruksjoner på morsmålet ditt som forklarer hvordan man setter sammen kinesiske tegn på riktig måte. Du forstår imidlertid ikke kinesisk (selv om du faktisk forstår kinesisk, tenk deg at du behandler tegn i et språk du ikke behersker). Du får nå kinesiske tegn som input og skaper kinesisk-språklig output ved å følge instruksjonene. Sett utenfra oppfører du deg helt likt noen som forstår kinesisk. Samtidig vet vi at dette ikke stemmer. Korrekt manipulering av kinesiske tegn er ikke det samme som å forstå kinesisk.

Searles poeng med tankeeksperimentet om det kinesiske rommet er at det du gjør når du manipulerer kinesiske tegn, i bunn og grunn er det samme som en datamaskin gjør med symboler i et programmeringsspråk. Akkurat som det er klart at en person som kun manipulerer kinesiske tegn, ikke forstår kinesisk, er det klart at en datamaskin som kun følger regler, ikke forstår og derfor ikke kan være intelligent i sterk forstand. Det er ikke nok at noe fungerer eller ser ut som forståelse. Virkelig forståelse, ifølge Searle, er et biologisk fenomen som krever bestemte nevralt prosesser i den menneskelige hjernen. Akkurat som man ikke kan produsere melk ved å kjøre en simulering av laktasjon (melkeutskillelse fra melkekjertler), bør vi ikke tro at sinnet kan forklares ved at det fungerer som en datamaskin (Searle, 1980, s. 424). Genuin intelligens trenger det riktige materielle grunnlaget, og det fins faktisk i hjernen. Så Searle tviler ikke på fysikalismens grunnleggende utsagn om at mentale