

Tittelen på kronikken

**Lars Bjørlykke Kristiansen Gorm Lundestad
Kristoffer Langerød**

1. november 2011

Ray Kurzweil tror at det en gang i mellom 2020 og 2070 vil bli laget en AI som består Turing testen og at i 2099 har maskiner overgått menneskene og fått samme rettigheter som mennesker. Etterhvert blir mennesket fredet og hyllet som skaperen av maskinene.

Fremtiden

Futuristen Raymond Kurzweil har skrevet en rekke bøker der han kommer med forutsigelser om teknologisk utvikling. Blant de mer kjente er forutsigelsen om at i 1998 ville en datamaskin kunne slå de beste menneskelige spillere i sjakk (Deep Blue slo Garry Kasparov i 1997), og i starten av det 21-århundret skulle internett få en eksplosiv økning i antall brukere og innhold, og vi skulle få datamaskiner i lomma som kan hente informasjon fra internett trådløst. Til sammen har idag kun en av hans 108 forutsigelser om teknologisk utvikling, fram til 2009, vist seg å være helt gal. Dette gir oss en pekepinn på at også hans utsagn om fremtiden, mest sansynlig har rot i virkeligheten.

I boken *The Singularity is Near*, antar han at det er mulig å lage en såkalt sterk kunstig intelligens, som vil si kunstig intelligens som er like god som, eller bedre enn menneskelig intelligens og at dette vil skje mellom år 2015 og 2045. I 2099 vil maskiner overgå menneskene «på alle mulige måter» og få samme rettigheter som mennesker. Etterhvert blir mennesket fredet og hyllet som skaperen av maskinene.

Reservedeler

Hvis et menneske mister en arm, ben eller nyre, kan ingen komme og si at vedkommende ikke lenger er et menneske, og det samme gjelder for mennesker som er født med ulike defekter". Og hvis man kan miste, eller fødes uten diverse organer som de fleste andre mennesker har, og likevel være en del av arten *homo sapiens*, da vil det vel ikke være noe i veien for å erstatte disse såkalte defektene med menneskeskapte hender, hår, ben osv.?

Vi har i veldig lang tid drevet med dette. I dag har briller blitt mer vanlig enn noen gang før, og i følge Vision Council of America bruker 75 prosent av alle amerikanere en form for synskorreksjon. Briller er selvfølgelig en mild form for å kompensere for defekter eller skavanker, men det er også eksempler på blinde som i en viss forstand har fått synet igjen, ved hjelp av et kamera, en datamaskin og kobling direkte inn i hjernen. Allerede i 2000 gav et slikt systemet, laget av Dr. Dobelle, syn nok til å navigere, til en tidligere blind mann. Og i 2002 hjalp dette systemet en bonde, mest kjent som Jens, muligheten til å kjøre bil. Hvor langt kan man gå i denne utbyttingsprosessen før man ikke lenger er et menneske? De fleste vil nok komme fram til at man kan bytte ut det aller meste i kroppen, bortsett fra hjernen, og fremdeles være et menneske. Men hva om man kunne bytte ut deler av, eller hele hjernen, og fremdeles være den samme personen?

Individ

La oss først som sist enkelt skille mellom en person og et menneske: Et menneske er et individ av arten *homo sapiens*, mens en person må ha egenskaper som selvbevissthet, mulighet til å utføre irrasjonelle handlinger, lære og erfare og så videre. Disse egenskapene, nevnt under person, er alle egenskaper som ikke går på fysiske ting, men heller væremåte eller egenskaper i sinnet. Så hvis disse egenskapene kan tenkes at en gang ikke så langt inne i fremtiden, som er nødvendige for å være en person, ikke er avhengige av en kropp, er det da så utenkelig at man kan gå til en bevissthetsbackup, for å lagre bevisstheten sin i ventemodus for sikker hetens skyld? Hva er egentlig forskjellen mellom minnene våre og det som er lagret på en harddisk til en datamaskin? Hvis vi så godtar at det er personen som er essensen i det å være menneske, og at alle de egenskapene som kreves for å være en person, kan dekkes av ikke-fysiske egenskaper. Da må man også betrakte et "individ" av en ikke-organisk karakter, som et individ, og til og med en person, hvis dette individet oppfyller kravene til å være en person. Alt annet ville være spesiesisme, og urim-

lig siden man da ikke legger vekt på kun relevante egenskaper. Hvorvidt det en gang i fremtiden, teknologisk vil være mulig, er et spørsmål ingen kan svare sikkert på. Men i boken *The Singularity is Near*, av den kjente futuristen Ray Kurzweil, antar han at det er mulig å lage en såkalt sterk kunstig intelligens, som vil si kunstig intelligens som er like god som, eller bedre menneskelig intelligens, mener Kurzweil at dette vil skje mellom år 2015 og 2045. Hvorvidt dette er et rimelig tidsestimat, eller i det hele tatt blir mulig, kan bare tiden vise - på samme måte som da Kurzweil i 1990 påsto at det skulle være mulig å koble seg trådløst på internett, og mange av hans andre prognoser.

Naturlig utvikling

Hvorvidt denne utviklen er naturlig kan diskuteres. Ut ifra dagens synspunkt vil de fleste si at det er rimlig unaturlig. Men hva hadde mennesker på 1600-tallet sagt om en boks med elektroder, operert inn i mennesket, som holdt hjertet i gang? Resultatet hadde nok vært at stakkaren som nevnte det, ble brent på bål som heks, eller drept på annen måte. Hva som er naturlig er betinget av tiden og samfunnet. Og siden den teknologiske utviklingen nok vil gå gradevis, vil det som oppfattes som naturlig, gradvis forandres og tøyes, enten vi vil det eller ikke. Spør man en person om fire hundre år, hva som er naturlig, vil man få helt andre svar enn hvis man spør mannen på gate i dag.

kunne en datamaskin/AI utviklet seg naturlig?

Kunstige rettigheter

Så kanskje det ikke er så lenge til en AI blir selvbevist. Eller hva hvis man kan overføre et menneskets tanker til en datamaskin.

noe mer nært Allerede idag har vi biler som kjører av seg selv og kun trenger en menneskelig sjåfør for å tilfredstille eksisterende lover. Det er ikke utenkelig at det om kort tid vil være lov for en bil å kjøre helt alene. Så la oss tenke oss det finnes biler som kan gjøre nettopp det. På vei hjem fra jobben har du kommet over noe du vil kjøpe, men som du ikke får med deg på toget. Da kan du starte en app på telefonen din og be den komme og hente deg.

På veien kommer bilen ut for en ulykke og ender med å ta liv av en person. Hvem er det som har ansvaret? Kanskje er det deg som eier av bilen, men du kunne aldri forutsett hva som kom til å skje på turen. Kanskje det er selskapet som produserte bilen, de burde kanskje ha forutsett hva som skulle skje. Eller vil det kanskje bli bilens feil.

Selvom det kanskje idag ville vært selskapet eller eierens skyld i dag så er det tenkelig at det i fremtiden vil komme AI'er som vil lære ting på egenhånd og selskapet ikke lengre har kontroll over hva den gjør.

Tenk deg en robot som tar lappen og består på samme måte som vi mennesker.

På et eller annet tidspunkt må vi innse at en AI har blitt et eget individ med egen vilje.

Når AI'er har blitt beviste og fått samme rettigheter som mennesker - hva da med alle industriroboter osv. Er det etisk akseptabelt å lage AI'er som er for dummetil å være selvbeviste? Det er jo kanskje mer praktisk med intelligente industriroboter så kanskje man lager AI'er som er intelligente nok til å være selvbeviste, men vi legger inn en sperre for at den skal gjøre som vi vil. Hva hvis man gjør det samme med mennesker? Lager mennesker med medfødt hjerneskade som kan brukes i industri og minerydding?

kanskje blir det å lage en AI i fremtiden mye av det samme som å bli forelder. Dersom man skal lage en AI må man gjøre det man kan for at AI'en skal følge lover og regler og at dersom man ikke gjør det ikke får lov til å lage AI'en. Etter dette vil det være de individuelle AI'ene som selv har ansvaret og skaperene kan kun få skylden dersom de burde ha lært opp AI'en bedre.

Konklusjon

Konklusjon: - AI og personer er det samme - AI kan bli selvbevist - Effekter på samfunnet - Ånskelig?