

Det er da også lett å finne dystre spådommer om hva KI vil innebære for menneskeheten dersom KI-ene blir smartere enn oss. Fysikeren Stephen Hawking mente for eksempel at selv om KI *kan* være det beste som har skjedd menneskeheten, så er det også stor sjanse for at det i stedet er det verste. Derfor må videreutviklingen av KI foregå veloverveid og strengt regulert, sa han (Cellan-Jones, 2014).

I tillegg til en lite underbygd antakelse om at smartere vesener enn oss vil ønske å utrydde oss, er premisset for Hawkings advarsel at menneskets intelligens er knyttet til den trege biologiske evolusjonen. Motsatsen er at den teknologiske evolusjonen som driver KI fremover, er langt raskere. Er det en holdbar forestilling?

Filosofen Alva Noë mener «nei». KI er informasjonsteknologi, og slik teknologi gjør i bunn og grunn bare regelstyrt informasjonsprosessering, om enn komplekst og lynraskt (Noë, 2014). Det gjelder enten reglene er programmert av mennesker, eller generert gjennom **maskinlæring**. Selv om ChatGPT imponerer oss, med velformede og meningsfulle setninger generert på rekordtid, er den i bunn og grunn bare drevet av statistiske muligheter for at språkmønstre som den har blitt trent på, passer inn i sammenhengen den brukes i (Falkum & Lison, 2023).

For Noë handler ikke naturlig intelligens, som den menneskelige, (kun) om statistikk og prosessering, men om kreativitet, mening og det å tilpasse omgivelsene til de behov, ønsker og preferanser vi har. Det er derfor en kvalitativ forskjell på naturlig og kunstig intelligens som ikke vil bli utlignet bare informasjon prosesseres kjappere og mer effektivt enn hva dagens KI-systemer kan gjøre.

Noë sammenligner KI med en klokke: Klokken *holder* tiden, men den *vet* ikke hva klokken er. En KI vil aldri være intelligent i seg selv, men når vi utvikler KI, så gjør vi oss *se/v* mer intelligente (Noë, 2014). KI er en måte vi skaper nye betingelser for våre liv, slik vi mennesker gjør gjennom all teknologi. Frykten for at menneskeheten skal bli etterlatt (og utryddet) av smartere KI, bygger slik sett på det som Noë mener er en misvisende tanke om at menneskelig intelligens kun sitter i hjernen og bare utvikles gjennom biologisk evolusjon.

Det meste av forskningen på og utviklingen av KI er imidlertid ikke opptatt av dystopier og filosofiske problemstillinger omkring intelligens-begrepet. KI er et praktisk felt, der man utvikler systemer som kan utføre bestemte oppgaver, på egen hånd eller i samband med mennesker. Det betyr dog ikke at bruken av og den voksende avhengigheten av dem er uproblematisk. Også KI-er er verdiladete: Deres tosidighet gjør at de ikke kan slippes løs i samfunnet uten etiske og politiske vurderinger.

Noen etiske utfordringer med kunstig intelligens

I 2016 lanserte Microsoft *Tay*, en maskinlærende chatbot. I løpet av noen få timer snudde *Tay* seg fra å være en hyggelig og imøtekommende samtalepartner til å bli en rasistisk og kvinnefiendtlig en. Årsaken var at Twitter-brukere fôret chatboten med bemerkninger i denne retningen, som den da lærte av (Hunt, 2016).

Tay ble utsatt for en bevisst kampanje, og utviklet seg som den gjorde fordi dataene den ble fôret med, var rasistiske og kvinnefiendtlige. Men også KI-systemer som lærer seg å gjenkjenne mønstre basert på tilfeldige data, kan utvikle seg i uønskede