Jonaslpe - iterasjon 1

KI - definisjonshistorie
Alan Turing publiserte en artikkel i 1950 der han stilte spørsmål rundt hvor vidt det var mulig å lage en maskin som tenkte selv, men begrepet kunstig intelligens (KI) ble først introdusert i 1956 av logiker og matematiker John McCarthy. Introduksjonen innebar optimistiske idéer og forestillinger KI aldri klarte å måle seg med, og historien har dermed vært preget av skuffelse og bølgedaler. Teknologien har enda langt igjen for å kunne måle seg med de optimistiske forventningene, men har samtidig kommet svært langt.

KI - 4 definisjoner
John McCarthy definerte i 2004 KI som "vitenskapen og konstruksjonen bak å lage intelligente maskiner, spesielt intelligente dataprogrammer. Det er relatert til den lignende oppgaven med å bruke datamaskiner for å forstå menneskelig intelligens, men AI trenger ikke å begrense seg til metoder som er biologisk observerbare." McCarthy har siden han introduserte begrepet i 1956 vært viktig for utviklingen av KI, og definisjonen bør anerkjennes. Tar man dog KI sin utvikling siden 2004 i betraktning kan det være nødvendig å kaste nytt lys på definisjonen.

Marvin Lee Minsky var en informatiker med særlig interesse for AI, og arbeidet ved MIT-fakultetet fra 1958 til sin død. Han startet i 1959 MIT Computer Science and Artificial Intelligence Labratory sammen med John McCarthy. I 1968 definerte han kunstig intelligens som "vitenskapen ved å få maskiner til å gjøre ting som ville kreve intelligens hvis de ble gjort av menn". Definisjonen deler likheter med McCarthy sin, og det kommer frem at de har jobbet tett sammen med fagfeltet.

Andreas Kaplan og Michael Haenlein, professorer med fokus på KI, definerte i 2018 KI som "et systems evne til å korrekt tolke eksterne data, å lære av slike data, og å bruke denne kunnskapen til å oppnå spesifikke mål og oppgaver gjennom fleksibel tilpasning". Definisjonen skiller seg fra de to andre definisjonene ved at den er mer spesifikk og tilpasset vår samtidens teknologi.

Feltets tverrfaglighet, bredde og hyppige utvikling gjør det utfordrende å stadfeste en definisjon. Elon Musk, en moderne pioner innenfor KI, har utfordringer med å definere KI, hvilket forsterker mitt inntrykk av hvor flytende begrepet er. Ved å bygge definisjonen min på en kombinasjon av samtiden og de nevnte definisjonene, vil jeg likevel definere KI som maskiner og teknologi som gjennom aktiv læring og analysering av store mengder data fleksibelt tilpasser seg over tid til å løse behov på et vis som fremstår så intelligent at det kunne vært gjort av et menneske.

KI - autonomi og automatikk
Med menneskelig autonomi menes individets evne til å være selvstyrende og moralsk uavhengig. Dersom man er autonom bestemmer man over seg selv og sine egne handlinger. En maskin anses derimot som autonom når den har evnen til å tilegne seg nok kunnskap om miljøet sitt til å være i stand til å gjennomføre oppgavene den er satt til å gjøre, selv når den nødvendige kunnskapen ikke var der fra start.

Til likhet med autonomi, fungerer noe som er automatisk helt av seg selv. Forskjellen ligger i mengde tilpasning, læring og beslutningstaking som er integrert i systemet. Autonome

systemer har færre begrensninger enn automatiske. De er dynamiske, lærer fort, og kan gjerne begi seg forbi sitt opprinnelige felt. Automatiske systemer begrenses av satte rammer og oppgaver, og valgene og oppgavene den gjennomfører er mer basert på forhåndsbestemt heuristikk.

Omtale av artikkel
I "Some Brief Thoughts on the Past and Future of Human-Robot Interaction" fra 2018 tar Kerstin Dautenhahn tanker og meninger om utviklingen av HRI hittil og veien videre. Hun adresserer blant annet suksessfaktorer, utfordringer, kjernebegreper og standarder. Hun stempler fagfeltet som en suksessfaktor de siste årene, og peker på årlige konferanser der fagfelt deler kunnskap, erfaringer og bekymringer på tvers av hverandre. Ironisk nok anses denne tverrfagligheten som en hovedutfordring, siden de ulike fagfeltenes forskjellige metoder og fokusområder gjør det vanskelig å felle fellesvurderinger over HRI. Tverrfagligheten har samtidig medført utfordringer tilknyttet begrepsstadfesting, for sentrale begreper som "robot". At eksperimentell psykologi er å anse som en gullstandard i HRI vekker bekymringer, siden det ikke tar høyde for virkelig interaksjoner mellom virkelige mennesker og virkelig komplekse roboter i virkelige omgivelser. Kvalitativ forskning med fokus på virkelige interaksjoner trekkes frem som nødvendig, til tross for at det er mer ressurskrevende. Til tross for utfordringene tverrfagligheten bringer, anser Dautenhahn dette som positivt for HRI, og hun understreker viktigheten av åpent sinn i møte med tilnærminger fra andre fagfelt.

Jeg har utfordringer med å forstå hvordan andre fagfelt er så verdifulle som det gis inntrykk av, og savner mer konkrete eksempler på informasjonsutveksling av verdi. Det er krevende å se for seg hvordan fagfolk uten teknologisk bakgrunn kan bidra med utviklingen av HRI. Med robot som flytende begrep savner jeg dessuten en klarere forståelse av hva som menes med robot og dermed også interaksjoner innen HRI. Dautenhahn nevner at den kvalitative forskningen bør bruke komplekse, intelligente og multitaskende systemer. Dette er forståelig nok, men når interaksjonen skal skje "face 2 face" maler hun et urealistisk og futuristisk bilde av en sci-fi robot som lever i tråd med fortidens forventninger til KI. Å forestille seg at denne interaksjonen skal skje i naturlige omgivelser blir dermed vanskelig. Ser jeg forbi dette er jeg ikke uenig med Dautenhahn.

AIBrain - moderne AI selskap
AIBrain ble stiftet i 2012, og har i mange år vært anerkjent som et av de beste AI-selskapene. Selskaper presenterer KI som et produkt som skal øke brukernes intelligens langs fire dimensjoner - kognitivt, sosialt, fysisk og emosjonelt. De viser til nevrale og kognitive vitenskaper for å støtte påstanden deres. Åpenhet og dokumentasjonen av teknologiens virkemåte gjør at selskapet fremstår profesjonelt og ærlig, istedenfor å presentere KI som "mystisk teknologi" der virkemåten er skjult bak kulissene. Landingssiden deres portretterer et bilde av en robot-hånd som speiler en menneskehånd på andre siden av et futuristisk grensesnitt. Bildet er åpent for tolkning, men representerer på mange måter den klassiske forståelsen av KI som menneskets erstatter. Utover dette presenteres KI på en profesjonell og realistisk måte.

I, Robot (2004) - film om samspill mellom AI og mennesker
Filmen finner sted i 2035 der menneskelignende roboter lever side om side med mennesker i samfunnet. Robotene fungerer er svært avanserte, og fungerer som brede og spesifikke oppgaveløsere. Samspillet mellom mennesker og roboter er tilsynelatende naturlig, og robotene fremstår som vennlige hjelpere uten følelser. Interaksjonen dem i mellom skjer

gjennom tale og kroppspråk, men robotene virker ikke avhengig av denne kommunikasjonen for å forstå hva de skal gjøre. De fremstår svært autonome i form av at de tar del i samfunnet og løser oppgaver på naturlig og logisk vis, slik et menneske ville gjort, men innenfor gitte etiske rammer. Filmen tar et vendepunkt når en teknofobisk etterforsker tar en robot i å begå et drap, og samspillet endres. Tilliten til robotene er svekket, og maktbalansen er snudd. Vendepunktet beskriver KI som har utviklet seg i mot sin hensikt og vendt seg mot mennesket, der roboten har beveget seg utenfor sine gitte rammer og handler ut ifra "fri vilje".

Robot - definisjonens oppstandelse
Ordet robot ble først introdusert for publikum av den tsjekkiske forfatteren Karel Capek i skuespillet R.U.R i 1920. Ordet stammer fra det slaviske ordet robota, som betyr "arbeid", eller "tvangsarbeid". Stykket viser en fabrikk som lager kunstige mennesker som ligner en premoderne forestilling av robot, slik man ser i filmen I, Robot.

Robot - 3 definisjoner
Dautenhahn anerkjenner at man innenfor HRI ikke har en felles definisjon for robot. Dette medfører på tvers av fagfelt og eksperter. Den offisielle definisjonen er midlertid "Actuated mechanism programmable in two or more aces with a deegree of autonomy, moving within its environment, to perform indented tasks." Definisjonen beskriver et maskineri som kan programmeres til å gjennomføre tiltenkt oppgave innenfor sitt miljø med en grad av autonomi. SNL definerer robot som "en maskin som kan utføre handlinger automatisk, og som ved hjelp av elektroniske «sanser» og «oppskrifter» på hva de skal gjøre (algoritmer) kan bearbeide inntrykk fra omgivelsene, som igjen påvirker hvordan disse oppgavene utføres". Her beskrives robot som mer automatisk enn autonomt, men til likhet med den offisielle definisjonen skal roboten være i stand til å dynamisk gjennomføre oppgaver basert på omgivelsene sine. Basert på de nevnte definisjonene vil jeg definere en robot som en autonom maskin som er i stand til å sanse omgivelsene sine, utføre beregninger for å ta beslutninger og utføre fysiske handlinger i den virkelige verden.

Relasjon mellom KI og roboter
KI og robotikk eksisterer både uavhengig og avhengig av hverandre. Ved å kombinere AI med robotikk får man en kunstig intelligent robot med et høyt nivå av autonomi som vil være i stand til å optimalisere oppgaver den er satt til å gjøre og "lære" de beste metodene. I kombinasjon vil KI vil på sett og vis fungere som robotens hjerne, og sensorene og de mekaniske delene som kroppen. Med en lite fastsatt definisjon av robot, og en flytende forståelse av KI er det midlertid utfordrende å sette grensen mellom hvor grensen mellom robot og KI går. Forskjellen blir dog tydelig ved at KI ikke interagerer med fysiske miljøer, f.eks kun gjennom algoritmer i dataprogrammer, samtidig som roboter gjennomfører fysiske oppgaver. Tar man min egen, samt den offisielle definisjonen av robot i betraktning, kommer det frem at fysiske oppgaver må gjennomføres for å kunne klassifisere noe som robot. Definisjonene betrakter samtidig roboter som autonome, til likhet med definisjonene av KI.

Roxo the FedEx SameDay bot - moderne fysisk robot
Roxo er FedEx sitt forslag på futuristisk logistikk. Roboten er designet for å levere pakker i en radius på tre til fem kilometere fra en forhandlers beliggenhet innen et døgn. Roboten skal kunne kjøre på fortau, sykkelfelt og veikanter, og beveger seg som en liten på fire hjul. Den har to støttehjul for å kunne manøvrere over objekter som fortauskanter, og sakker farten før en endring i bevegelse, slik at bevegelsene dens gir et visuelt naturlig inntrykk. Roboten kommuniserer gjennom en stor og tydelig skjerm, samtidig som den har en begrenset talefunksjon.

UU - definisjoner
Universell utforming (UU) er nødvendig for inkluderende design i flere fagfelt, slik som IKT og byggebransjen. Forskriften om UU innen KTI ble oppdatert i 2022, og definerer UU som "Utforming eller tilrettelegging av hovedløsningen i informasjons- og kommunikasjonsteknologi slik at virksomhetens alminnelige funksjon kan benyttes av flest mulig". Definisjonen vektlegger at flest mulig, altså også de med bruksvansker slik som mennesker med kognitive utfordringer, skal kunne benytte seg av virksomhetens digitale hovedløsning. Definisjonen deler mye med definisjonen i diskriminerings- og tilgjengelighetsloven som ble revidert i 2013, der den tydeligste forskjellen er at denne definisjonen også inkluderer fysiske forhold: «Med universell utforming menes utforming eller tilrettelegging av hovedløsningen i de fysiske forholdene, herunder informasjons- og kommunikasjonsteknologi (IKT), slik at virksomhetens alminnelige funksjon kan benyttes av flest mulig». Ronald Lawrence Mace var en av grunnleggerne bak UU, og definerte konseptet som "design som kan brukes av alle mennesker, i størst mulig grad, uten behov for tilpasning eller spesialisert design". Til tross for at definisjonen er over 25 år gammel, er den fremdeles relativt presis og dagsaktuell, og det viser hvilken felles forståelse man har av UU både på tvers av og innad i fagfelt. I forhold til inkludering handler UU om å designe for at flest mulig skal kunne bruke en tjeneste, til tross for at de har ulike forutsetninger.

KI - menneskelighet og potensiale for inkludering
Siden KI ble introdusert har offentligheten hatt optimistiske forventninger til utviklingen. KI har kommet langt, men klarer enda ikke gjenspeile mennesker på en naturlig måte. Vår evne til persepsjon, bevegelse, og følelser er faktorer som skiller mennesker fra dyr - og teknologi. Persepsjon er sanseinntrykk og den påfølgende tolkningen av disse, og skjer i to trinn: gjennom stimuli av ett eller flere sanseorgan, og at stimuleringen tolkes og resulterer i en opplevelse. Vi har ulike forutsetninger for persepsjon, basert på tidligere inntrykk, og tilgjengelige sanser. En KI, eller en robot, vil også ha ulike forutsetninger for persepsjon. Denne vil da baseres på hvilke komponenter den bruker til å sanse og analysere omgivelsene sine, og hva den har lært. Dagens teknologi gjør at KI har potensiale for enda flere og sterkere sanser enn oss. KI bruker dyp læring for å tolke mønstre i data for å ta en endelig avgjørelse, hvilket på mange måter ligner menneskelig beslutningspsykologi. KI bruker samtidig innkodete standardinstruksjoner for bearbeiding av logikk og begrunnelse, hvilket ligner menneskelig rasjonell intelligens. Det kan virke som KI sine forutsetninger for å oppnå menneskelig persepsjon er gode, om ikke bedre enn våre egne, men faktumet at tolkningen av sanseinntrykkene deres ikke resulterer i en opplevelse er avgjørende for hvorfor de likevel ikke har evne til persepsjon.

Det finnes ingen KI som er i stand til å bevege seg slik som mennesket gjør. At bevegelsene våre er komplekse gjennom konstant kalibrering av balanse, fart, mimikk og kroppspråk gjør det vanskelig å kopiere bevegelsene våre. Det finnes menneskelignende roboter, slik som Ameca the humanoid, som klarer å etterligne menneskelig atferd, men som røpes av unaturlig ansiktsmimikk, og en manglende helhetlig kroppsbevegelse. Roboten vil heller ikke være i stand til så mye mer enn å kopiere atferd og å gjennomføre simple oppgaver, så de fysiske begrensningene hindrer den fra i å f.eks spille fotball. KI har ingen evne til å føle, kun evnen til å vite hvordan den skal respondere til sin egen tolkning av dataen den registrerer. Våre følelser er basert på hormoner og er ikke noe vi nødvendigvis har kontroll over, i motsetning til KI sin frie tolkning av binære tall. Dersom man vender blikket mot vanligere KI, som ikke er skapt for å fysisk etterligne et menneske, slik som en smarthøyttaler, blir den manglende tilstedeværelsen av de menneskelige egenskapene enda tydeligere. Til tross for manglende

menneskelige egenskaper, kan KI ha potensiale for inkludering. Et godt eksempel er gjennom appen "seeing AI", som er til for å assistere blinde. Appen har flere moduser, og forteller høyt hva som oppfattes foran skjermen, slik at den blinde vet hva den står ovenfor. Smarttelefoner med talegjenkjenning og bildegjenkjenning forenkler også hverdagen til utallige med nedsatt funksjonsevne eller nedsatt syn gjennom KI.

KI- inkluderende eller ekskluderende?

KI har både potensiale for inkludering og ekskludering. Funksjoner som talegjenkjenning og bildegjenkjenning for blinde er tydelige eksempler på hvordan KI kan inkludere de med særskilte behov. Dette er midlertid KI som er designet for spesifikk inkludering. Ved design for massene er KI svært mottagelig for å ekskludere siden den lærer basert på data fra massene. Ofte brukes normalfordeling i behandlingen av datasettene, og dette fører til at de som avviker fra normalbruk blir utelatt. Samtidig har KI potensiale for rasisme og sexisme. Dette kan både være basert på datasett KI-en selv har plukket opp som viser korrelasjon mellom kjønn eller rase og verdien KI-en er interessert i, eller at feil i algoritmene får KI-en til å diskriminere basert på rammene den er gitt. Eksempel på rasisme integrert i KI ble i 2019 funnet i det amerikanske helsevesenet på et system brukt på over 200 millioner mennesker for å forutsi hvilke pasienter som hadde behov for medisinsk behandling. Algoritmen prioriterte hvite pasienter fremfor svart pasienter basert på helsekostnadshistorie som viste seg å som regel være høyere blant hvite mennesker. Selv om rase ikke var en variabel i algoritmen prioriterte algoritmen de med høyest helsekostnadshistorie, som altså viste seg å korrelere med rase. Det ble i 2015 avduket feil i Amazon sitt ansettelsessystem som til likhet skilte mellom menn og kvinner basert på antall søknader som var innsendt de siste 10 årene. Systemet var opptrent basert på søknadene som var sendt inn det siste tiåret, og siden majoriteten av søknadene var fra menn var systemet opptrent til å favorisere menn.

KI - kan maskiner forstå?

Forståelighet er som en av fire grunnprinsipper en grunnpilar for tilgjengelighet på nett. Prinsipper dreier seg i all hovedsak om at brukere må være i stand til å forstå informasjonen så vel som hvordan grensesnittet opereres. Forståelse på begge ender av grensesnittet er viktig, og setter skillet mellom å forstå og forståelse. Brukeren må forstå både informasjonen som gis og grensesnittet som tilbys, men tjenesten må også være i stand til å forstå brukeren, også dersom brukeren velger å gi input gjennom utradisjonelle metoder. Hvor vidt maskiner har kapasitet for forståelse kan diskuteres og er i stor grad et definisjonsspørsmål. Hva er en maskin og hva er forståelse? Maskiner kommer i alle størrelser og varianter, og det gjør også forståelse. To individer kan ha samme opplevelse uten å dele den samme forståelsen av opplevelsen. Man har enten forståelse for eller av noe. Om man har forståelse for noe har man empati, og forståelse for personer, følelser eller folks situasjoner, og gjerne ved å tidligere ha vært i lignende situasjon selv. I dagligtale brukes uttrykk som å "sette seg i andres sko eller fotspor", og dette er noe en maskin vil ha utfordringer med å gjøre, siden den naturligvis ikke har levd et liv på linje med oss selv. Ved å ha forståelse av noe har man forståelse av kunnskap og oppfatninger, slik som grammatikk eller geografi. En maskin har absolutt kapasitet for kunnskap, men hvor vidt den faktisk forstår kunnskapen sår jeg tvil over.

Retningslinjer for menneske-KI interaksjon

Prinsipp 13, "learn from user behaviour", i Microsoft sine retningslinjer for menneske-AI interaksjon låter "Personalize the users experience by learning from their actions over time", og dreier seg om å gi brukeren skreddersydde opplevelser basert på tidligere bruk. Brukerens handlinger under bruk gjenspeiler ikke nødvendigvis brukerens behov og ønsker. Bruken kan ha vært basert på at brukeren var nysgjerrig, distrahert, klikket feil, eller var i en annen

uvanlig situasjon som gjorde at bruken skilte seg fra normalbruken. Brukeren kan dermed skape sin egen bias basert på et usedvanlig bruksmønster som systemet ikke klarer å skille fra normal bruk. Et typisk eksempel på dette er når man får opp haugevis av reklame på nett for ting man har klikket på ved uhell, eller kun klikket på av nysgjerrighet. Slik atferd redegjør ikke prinsippets beskrivelse for, så et logisk alternativt eksempel for prinsippet kunne vært "Personalized user experience based on user preferences and repetitive actions over time".

Shneidermans åtte retningslinjer innenfor HCI er skapt for å hjelpe designere med å løse problemer, og har flere likheter og forskjeller med Microsoft sine menneske-KI interaksjonsretningslinjer. Begge settene med retningslinjer setter brukbarhet og brukerens forståelse i sentrum. Eksempelvis understreker begge settene at informativ feedback er nødvendig for at brukeren skal forstå hva som skjer bak skjermen, eller at man enkelt skal kunne avslutte en prosess. Til forskjell fra Microsoft sine retningslinjer fokuserer HCI-retningslinjenen på consistency. Det er utfordrende for KI å være consistent når teknologien bygger på stadig utvikling. Microsoft sin imøtekommelse med utfordringen adresseres dog av prinsipp 14, som foreslår at oppdateringer og tilpasninger må skje på varsomt vis, samtidig som settet med retningslinjer understreker viktigheten av å unngå å forvirre brukeren ved å alltid informere om hvorfor systemer oppfører seg som det gjør.

Kilder:

- Dautenhahn, K. (2018). Some Brief Thoughts on the Past and Future of Human-Robot Interaction. ACM Transactions on Human-robotic Interaction, 7(1), 1-3.
- Bratteteig, T., & Verne, G. (2018). Does AI make PD obsolete? ACM International Conference Proceeding Series, 2, 1-5.
- Grudin, J. (2009). AI and HCI: Two fields divided by a common focus. The AI Magazine, 30(4), 48-57.
- Christoforou, E., & Müller, A. (2015). R.U.R. Revisited: Perspectives and Reflections on Modern Robotics. International Journal of Social Robotics, 8(2), 237-246.