Individuell oppgave

Høst 2020 | Siljehlu | 3. iterasjon

Innhold

1.1 Concepts, definition and history of AI and interaction with AI

Alan Turing skrev i *the London Times* i 1949 at «I do not see why [the computer] should not enter any one of the fields normally covered by human intellect, and eventually compete on equal terms" (Grudin, 2009, s. 49). Seks år senere, i 1956, dukket begrepet *artificial intelligence,* eller kunstig intelligens, opp for første gang. John McCarthy arrangerte en konferanse med navnet «The Dartmouth summer research project on artificial intelligence», hvor han inviterte forskere fra ulike felt for å diskutere og utvikle konseptene rundt «tenkende» maskiner (Marr, 2018).

Kunstig intelligens har siden den gang blitt definert mange ganger. Britannica (2020) definerer AI som "the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings". Amazon (u. å.) definerer AI som "the field of computer science dedicated to solving cognitive problems commonly associated with human intelligence, such as learning, problem solving, and pattern recognition". En tredje definisjon er fra Russell og Nordvig i 2010, og de definerer kunstig intelligens som "a subfield of computer science aimed at specifying and making computer systems that mimic human intelligence or express rational behaviour, in the sense that the task would require intelligence if executed by a human" (Bratteteig og Verne, 2018, s. 1-2). Definisjonene vektlegger forskjellige aspekter ved kunstig intelligens, men de har til felles at det handler om at et system, en maskin eller robot klarer å utføre handlinger eller prosesser som vanligvis er forbeholdt mennesker. Som Amazon nevner i sin definisjon, kan dette være i form av læring, problemløsning og mønstergjenkjenning. Det er altså et forsøk på å gjenskape en form for menneskelig kognisjon gjennom et system eller en maskin.

Basert på dette har jeg laget min egen definisjon: «Kunstig intelligens er når et system oppfatter informasjon, og klarer å bruke denne informasjonen senere i en kontekst for å ta beslutninger, kommunisere med brukere eller å løse problemer». I denne definisjonen er læring et sentralt konsept fordi intelligens uten læring er meningsløst. Uten evnen til å bruke informasjon og de erfaringene vi har fra tidligere er vi alle like ubrukelige, og jeg mener det er kanskje det som skiller et system med kunstig intelligens, og et system som er programmert til å utføre en rekke spesifikke oppgaver.

Itera er et konsulentfirma som blant annet jobber med AI i Norge. Kunstig intelligens blir presentert som at det vil være en av de ledende teknologiene i fremtiden, og Itera ser på dette

som en mulighet til å skape innovative løsninger for sine kunder. På denne måten blir AI sett på som noe som gjør sluttprodukter bedre og mer attraktivt.

I filmen Blade Runner (1982) portretteres interaksjonen mellom menneske og kunstig intelligens som problematisk. På den ene siden har masseproduksjonen av menneskelignende androider med kunstig intelligens, replicants, tillatt menneskene å kolonisere galaksene gjennom å gi tilgang på et lydig, effektivt og intelligent alternativ til menneskelig arbeidskraft. På den andre siden blir kunstig intelligens i filmens verden møtt med mistanke og avsky. Androidene som har drevet menneskets kolonisering av stjernene blir forbudt å returnere til jorden, og det finnes egne seksjoner av politiet som blir opprettet for å luke ut og likvidere androider som skjuler seg i det menneskelige samfunnet. Et av temaene i filmen er hvorvidt disse androidene også kan sees på som personer, til tross for at de er skapt i fabrikker. Selv om AI-teknologien i Blade Runner kanskje er ganske fjern fra hva kunstig intelligens i dag er tror jeg den belyser en grunnleggende bekymring vi mennesker kanskje har for å dele vårt monopol som tenkende, rasjonelle vesen med noen andre.

1.2 Robots and AI systems

Ordet robot er egentlig et Tjekkisk låneord som betyr tvangsarbeider. Det ble først brukt om mekaniske mennesker i boken *Rossums Universale Robotter* skrevet av Karel Čapek i 1920. Robot stammer ellers fra *robota* som betyr arbeid (de Caprona, 2013, s. 1215).

I 1979 definerte the Robot Institute of America robot som "*a reprogrammable, multifunctional manipulator designed to move materials, parts, tools, or specialized devices through various programmed motions for the performance of a variety of tasks*" (Thrun, 2011, s. 11). Denne definisjonen er veldig mekanisk, og ser på roboten som maskin som programmers for å utføre spesifikke oppgaver. Den gir ingen rom for autonomi. I kontrast til denne definerer the Merriam Webster's collegiate dictionary (1993) en robot som "*an automatic device that performs functions normally ascribed to humans or a machine in the form of a human*" (ibid). Her går de mer inn på det menneskelige og intelligente aspektet ved roboter. Atferd er noe som skjer automatisk, og den utfører oppgaver som i utgangspunktet man tenker at det krever et menneske til å utføre.

Disse to definisjonene er veldig forskjellige, men jeg mener begge har noe ved seg som gjør dem nyttige. Begge beskriver en robot som noe fysisk. Bevegelsesaspektet i den første definisjonen er interessant, selv om den i utgangspunktet bare beskriver en automatisert maskin. Den andre definisjonen sier at roboter gjør oppgaver som er forbeholdt mennesker.

Min definisjon vil da være følgende: «En robot er et fysisk objekt som egenhendig kan utføre en eller flere oppgaver».

Roboter kan inneholde større eller mindre grad av kunstig intelligens Tar man utgangspunkt i definisjonen fra the Robot Institute of America så kan en robot fungere fint uten noen form for kunstig intelligens, siden den kan være ferdig programmert til å gjøre sine oppgaver, og er ikke forventet at den forbedrer seg uten at koden forbedres. Det holder at den kan bli programmert til å utføre sine oppgaver uten at mennesker må hjelpe den. Et typisk eksempel kan være en robot som er en del av en produksjonskjede. Roboter som inneholder større grad av kunstig intelligens vil være mer avanserte, og kunne utføre andre oppgaver enn de som er uten. Den vil kunne lære seg nye ting, og kanskje utføre oppgaver bedre etter hvert som den får erfaring. Ser man på forholdet mellom roboter og kunstig intelligens, så er kunstig intelligens noe en robot kan ha mer eller mindre av.

Pepper er en robot utviklet av SoftBank. Ifølge hjemmesidene deres er hun den første menneskeliknende robot som kan kjenne igjen ansikter og enkle emosjoner hos mennesker. Hun er en assistent som kan brukes til å ønske velkommen og informere gjester på en innovativ måte. Man kan interagere med henne gjennom tale eller touch-skjerm, og hun kan kjenne igjen 15 ulike språk. Pepper sine bevegelser er menneskeliknende. Hun har ikke ben, men to armer og et hode i en menneskeliknende figur som gestikulerer mens hun prater. I et intervju med Pepper spør journalisten: «Should I be afraid of you?". Responsen fra Pepper var: "Have you seen my Instagram? I'm just plain cute!" (Tech Insider, 2018, 0:01). Både igjennom utforming og interaksjon utstråler roboten Pepper vennlighet ovenfor brukeren.

1.3 Universal Design and AI systems

Universell utforming er blitt definert av Ron Mace i 1985 som "design that's usable by all people, to the greatest extent possible, without the need for adaptation or specialized design" (The Universal Design Project, 2020). Denne definisjonen beskriver at universell utforming handler om å finne løsninger som fungerer for flest mulig mennesker, uten behovet for at noen trenger særegne løsninger for å kunne delta på lik linje med alle andre. På denne måten kan man inkludere flest mulig til å delta, uten fare for å bli stigmatisert eller sett på som annerledes fordi man ikke klarer det samme som alle andre. Det handler om design for alle.

Kunstig intelligens har utviklet seg med tanke på å gjøre ytterligere ting tilgjengelig for mennesker som tidligere ikke har kunnet delta på samme måte. Et eksempel her er selvkjørende biler. Disse gjør det mulig for mennesker som av ulike årsaker ikke kan kjøre bil

til å være mer selvstendig og mindre avhengige av andre for å komme seg rundt. Et annet eksempel er smarte høreapparat som forbedrer lydkvalitet og tilpasser seg brukeren slik at den får en bedre opplevelse enn om kunstig intelligens ikke var til stede. På denne måten kan kunstig intelligens bidra til at flere får en enklere og friere hverdag, og at personer med ulike nedsettelser kan delta på samme måte som de uten nedsettelser.

Kunstig intelligens sin største svakhet er at maskinen bare kan lære av den dataen man gir den. Det vil si at systematiske bias som eksisterer i samfunnet i dag vil videreføres gjennom maskinlæring. Et klassisk eksempel her er ansettelsesprosesser. Dersom de fleste ledere er hvite menn på 50 år, så vil kanskje systemet lære seg at det er denne profilen den skal se etter når den ser etter en ny leder. På denne måten kan kunstig intelligens være med på å forsterke eksisterende skjevheter. Men, til syvende og sist er det mennesker som skal programmere hvordan AI-et fungerer, og dette gir oss muligheter til å rette opp i disse skjevhetene. Så lenge man er bevisst på hvilke biaser som er en del av samfunnet i dag og hvilke vi står ovenfor i fremtiden, så kan vi utvikle kunstig intelligens som omgår dette. Det krever kunnskap i bunn, men på den måten kan kunstig intelligens være med på å skape et rausere, mer inkluderende samfunn som gir like muligheter for alle.

Når det kommer til begrepene «å forstå» og «forståelse» så handler det om de kognitive prosessene mennesker gjør rundt noe. I løpet av livet skaper mennesker mange ulike knagger å henge ting på, og avhengig av konteksten kan samme tingen bety noe helt annet. En datamaskin kan forstå en input innenfor de rammene som vi har gitt dem. Den vil kunne kjenne det igjen og gjøre operasjoner fortere enn et menneske kan gjøre det. Likevel vil den ikke, per i dag, ha forståelse for de kulturelle og sosiale konsteksetene begrepet innebærer.

1.4 Guideline for Human-AI interaction

Guideline 4: Show contextually relevant information
Denne retningslinjen handler om at systemet skal vise informasjon som er relevant i den konteksten det er i. Et eksempel på dette kan være dersom jeg har søkt på et telefonnummer i telefonkatalogen. Dersom i neste steg åpner Google Maps, så vises adressen tilhørende det telefonnummeret dersom jeg begynner å skrive inn en adresse som likner.

Dersom vi sammenlikner Schneiderman's åtte gyldne regler med Microsoft sine Guidelines for Human-AI Interaction, så er det retningslinjer som går på to vidt forskjellige aspekter av design. Schneiderman's gyldne regler handler om hvordan man skal utforme brukergrensesnittet, mens Microsoft sine retningslinjer handler om interaksjonen mellom

mennesker og AI. Microsoft deler sine retningslinjer inn i 4 forskjellige faser av interaksjonen (før og under interaksjon, når systemet gjør feil, og over tid) mens Schneiderman's regler er mer statiske og generelle for hvordan utformingen bør være. Likevel så har de noen likheter. Begge har for eksempel fokus på at brukeren får informativ respons på interaksjonen, at brukeren kan unngå feil eller reversere handlingene sine. En del av dem går inn på de samme aspektene ved designet, men de kommer med et helt ulikt mål.

2.1 Characteristics of AI-infused systems

Følstad (2020, slide 16) identifiserer fire karakteristikker ved et AI-basert system. Den første karakteristikken er læring. Læring er viktig for at systemet skal klare å ta imot informasjon, bearbeide denne, og deretter utføre oppgaver bedre over tid. Dette gjør at disse systemene er avhengige av kontekst og mulig sårbare for at små endringer i input kan påvirke prestasjonen videre. Dette har sammenheng med den neste karakteristikken som er at AI-systemer forbedrer seg. Når systemene kan mere, så utfører de oppgavene bedre. Dette impliserer også at systemer gjør feil, og systemet bør være designet på en måte som effektivt korrigerer disse. Kocielnik et al. (2019) sin studie viser hvordan brukernes forventning til hvilke feil som kan oppstå er viktig for oppfatningen av hvor nøyaktig systemet er i bruk. Mer om dette i oppgave 2.2. En tredje karakteristikk er «black box». I et AI er det mange prosesser som skjer, og man som bruker forstår ikke alt som skjer. Derfor kan det være vanskelig å forstå hvorfor man får en output, siden man ikke vet hvilken informasjon AI-et har, og hvordan AI-et prosesserer denne. Yang et al. (2020) beskriver to årsaker til hvorfor det kan være vanskelig å designe for brukeropplevelser, nettopp det at er usikkerhet knyttet til hva AI-et kan gjøre. I tillegg er AI-et komplekst, og outputen kan vises på en enkel eller kompleks måte. Disse er noen av tingene som skjer i black-boksen. Den siste karakteristikken ved AI-baserte systemer er at de er drevet av store datasett. Denne dataen oppnår de gjennom interaksjon med brukere og systemer.

Et eksempel på et AI-basert system er videotjenesten YouTube. YouTube er bygget opp av algoritmer som lærer seg hvilket innhold du som bruker foretrekker, og gir deg anbefalinger basert på disse. Når disse igjen analyserer hvordan du mottar anbefalingene, så vil den bli mer og mer nøyaktig å finne innhold som er interessant for deg. Et eksempel på en konsekvens av dette kan være at en bruker får de politiske synene sine forsterket ved hjelp av algoritmen dersom brukeren bare får opp videoer og innhold som bekrefter det synet man allerede har. Algoritmene fortsetter å servere brukeren det samme budskapet, og med mindre brukeren selv oppsøker motstridende meninger kan det gi et inntrykk av at ens syn er det eneste fornuftige alternativet. Her har alle de fire karakteristikkene innvirkning på hvordan denne selvforsterkende spiralen fungerer: Systemet lærer seg hva du liker, og forbedrer seg og forslagene sine. Brukeren gir fra seg data som systemet tar med seg videre og forer inn i AI-et. Det hele skjer i en black box hvor brukeren ikke nødvendigvis har innsyn i hva som lagres, hvor mye som lagres, og når det lagrer det. Dette er selvsagt bare en mulig konsekvens. På den positive siden vil brukerne i større grad oppleve at de mottar forslag på innhold de selv er

interessert i. Mange vil nok verdsette dette og synes dette er en del av opplevelsen som tjenesten tilbyr.

2.2 Human-AI interaction design

Amershi et al. (2019) presenterer 18 designretningslinjer for menneske-AI-interaksjon. Disse retningslinjene er ment som en ressurs for å jobbe med systemer og tjenester som inneholder kunstig intelligens, og for videre forskning på feltet. Artikkelen beskriver metoden for hvordan de har kommet fram til disse; fase 1 var en innsamlingsfase hvor de samlet designanbefalinger for kunstig intelligens fra 1. anmeldelse av produkter fra industrien, 2. nylig publiserte artikler om AI-design, og 3. relevante forskningsartikler om AI-design. Fase 2 bestod av en intern evaluering av funnene fra fase 1. I fase 3 benyttet de seg av en brukersstudie hvor de samlet personer som jobber med HCI til vanlig, for å 1. se om de forstod retningslinjene i praksis, og 2. få tilbakemelding på tydeligheten av dem. Til slutt i fase 4 hadde de en ekspertevaluering av endringene de hadde gjort fra brukerstudien.

Kocielnik et al. (2019) beskriver en studie hvor de viser viktigheten av det å sette de rette forventingene til et AI-system for at brukeren skal ta det imot på riktig måte. De bruker Scheduling Assistant, et AI-system som automatisk oppfatter møteforespørsler i eposter. I studien utforsker de to versjoner av samme system med samme treffsikkerhet, og viser at forskjellig fokus kan endre den subjektive opplevelsen av treffsikkerhet og aksept av programmet. De foreslår tre teknikker for å legge grunnlaget for forventninger: 1. en *accuracy indicator* som eksplisitt sier hvor nøyaktig systemet er, 2. eksempelbasert forklaring, som prøver å øke forståelsen til brukeren, og 3. *performance control* som tillater brukeren å direkte tilpasse prestasjonen. Studien viser hvordan disse teknikkene bevarer brukertilfredsheten og aksepten for at et system ikke er perfekt. Studien viser også at AI-systemer som unngår falske positive feil kan føre til at brukerne føler at systemet er mindre treffsikkert enn det det er, og lavere aksept enn systemer som prøvde å unngå false negativer – til tross for at de var like treffsikre.

Amershi et al. (2019) sine 18 designretningslinjer for menneske-AI-interaksjon kan brukes til å analysere eksempelet med YouTube i oppgaven over. G13 (guideline 13) som omhandler læring fra brukeratferd er veldig relevant med tanke på det tidligere eksemplet. YouTube har virkelig satset på å personliggjøre brukeropplevelsen med deres tjenester ved å analysere alle brukerinput og brukerdata for å lære mer om hvordan systemet kan gi en bedre, mer tilpasset

opplevelse for brukeren. En annen retningslinje som også er interessant i denne sammenhengen kan være G8 om å støtte effektiv avfeiing. Hvis du som bruker for eksempel får opp en innholdsskaper eller video du ikke er interessert i å se mer av, kan du enkelt gi beskjed til systemet om at du ikke lengre er interessert i den typen videoer, eller at du ikke vil se flere forslag fra den skaperen. YouTube ser ut til å ha et bevisst forhold til hvordan AI-et sitt fungerer ovenfor brukerne, og det kan tilsynelatende se ut som om at de har prøvd å forbedre disse punktene så mye som mulig. Det de derimot ikke kan kontrollere er hvordan for eksempel det at systemet lærer brukeren å kjenne og hvordan brukeren tar imot at systemet gjør dette. For noen kan det oppfattes ubehagelig at tjenester lagrer informasjon og klarer å danne et godt bilde av hvem man er som bruker. Et kompromiss her kunne vært å bruke tjenesten uten de fordelene som de personlige tilpasningene kan tilby, til fordel for at man som bruker ikke blir husket.

2.3 Chatbots / Conversational user interfaces

Følstad og Brandtzæg (2017) beskriver 3 implikasjoner for HCI; Samtaler som designobjekt, behovet for å endre fokus fra brukergrensesnitt til tjenestedesign, og behovet for å designe nettverk for både mennesker og bots.

Samtaler som designobjekt: Design av chatbotter skiller seg fra visuelt design med tanke på at hovedfokuset ligger på interaksjonen mellom bruker og chatbot, og mekanismene for å designe samtaleflyten. Design for brukbarhet er her viktig ettersom mye av funksjonaliteten til en chatbot kan være skjult ovenfor brukeren. Måten å designe denne type interaksjon blir annerledes enn grafisk design alene, fordi man må designe for selve interaksjonen og gi brukene forslag og forklare hva den kan brukes til, og legge til rette for å utnytte de funksjonene og mulighetene som er til stede. Følstad og Brandtzæg (2017) argumenterer for at man burde gå fra å se på design som en forklarende oppgave, til at det skal bli en fortolkende oppgave når det kommer til design av samtaler. Det vil si å gå fra å forklare for brukeren hvilket innhold som er tilgjengelig, til å forstå hva brukeren trenger og forstå hvordan man best kan støtte dette behovet (s. 41).

Behovet for å endre fokus fra brukergrensesnitt til tjenestedesign: Følstad og Brandtzæg (2017) argumenterer for at det er mulig at i fremtiden vil all interaksjon med forskjellige nettsider og tjenester mulig gå over de sammen samtaletrådene, og det at det hele vil smelte

sammen. Med tanke på at fokuset i HCI til i dag i stor grad har hatt søkelys på enkeltbrukeren, mener de at det kan være fornuftig å rette mer fokus på tjenestene som en helhet.

Behovet for å designe interaksjon nettverk for både mennesker og bots: Denne implikasjonen tar for seg utfordringen å designe for flere agenter. HCI-design i dag tar stort sett bare for seg en bruker og en tjeneste. Det i seg selv er greit, men det betyr ikke at de ikke finnes flere måter å gjøre det på, og det finnes større sosiotekniske systemer som krever at man designer for flere agenter enn en en-til-en-interaksjon. Følstad og Brandtzæg (2017) eksemplifiserer dette med Microsoft sin chatbot Tay, som etter kort tid ble manipulert og lært opp av en rekke brukere til å bli, mildt sagt, ytterliggående.

Den første kategorien av retningslinjer mellom mennesker og kunstig intelligens som Amershi et al. (2019) presenterer går på de tingene man kan designe før selve interaksjonen skjer. G1 handler om at man gjør det klart ovenfor brukeren hva systemet kan gjøre. Dette samsvarer med implikasjonen om samtaler som designobjekt, hvor man må designe for samtalen, og på en eller annen måte klare å indikere ovenfor brukeren hva den er kapabel til. G2 går ut på at systemet skal gi en indikasjon på hvor godt det kan utføre de tingene den kan gjøre. Dette handler om at systemet skal hjelpe brukeren å forstå hvor ofte det gjør feil, slik at brukeren kan ha en realistisk forventning til hvor godt systemet presterer. Kocielnik et al. (2019) sin studie viser viktigheten av at systemene gir de rette forventningene til brukerne for at opplevelsen til brukerne skal oppleve at systemet fungerer så godt som det lover. Dette er igjen relevant for den utfordringen det er å kommunisere eller med å instruere brukerne.

3.1 Human-AI collaboration

Phillips et al. (2016) beskriver og eksemplifiserer hvordan team bestående av mennesker og dyr samarbeider om å utføre oppgaver, og hvordan dette har inspirert til å utvikle roboter som skal sees på som et team-medlem heller enn et verktøy. De beskriver hvordan de har overført kunnskap om effektive menneske-dyr-team, og hvordan dette har fungert til tross begrensninger når det kommer til selvstendighet og kommunikasjon. De utvikler en taksonomi hvor dyr enten erstatter (*replace*), multipliserer (*multiply*) eller utvider (*augment/extend*) menneskelige fysiske, emosjonelle eller kognitive ferdigheter, og overfører disse til utviklingen av roboter.

Oppgaven vil videre ta for seg to eksempler på roboter som har bidratt med å utføre oppgaver som dyr tidligere har utført. Eksemplene vil ta for seg Big Dog, en robot som erstatter fysisk kraft, og Paro, en robot som tilbyr emosjonell omsorg.

3.1.1 Eksempel 1: Big Dog

Big Dog er en robot utviklet av Boston Dynamics som skal bære bagasje i en militær setting, for å avlaste vekt fra soldater. Selve designet skal forestille en stor hund eller et lite muldyr, og tanken er at den skal navigere gjennom vanskelig og usikkert terreng (Phillips et al., 2016, s. 104). Ser vi tilbake på taksonomien til Phillips et al., så er hensikten til Big Dog å erstatte fysiske ferdigheter som tidligere har vært utført av levende dyr som esler, muldyr eller andre trekkdyr.

Big Dog er en robot som er selvgående i stor grad. Dersom vi ser på Big Dog opp mot Sheridan og Verplank sine nivåer at autonomi kan den plasseres rundt 7-8. Den utfører sine oppgaver automatisk, og tilpasser seg terrenget den beveger seg i. Ser vi på Big Dog opp mot Shneiderman (2020) sin todimensjonale akse vil jeg argumentere for at den scorer høyt på dataautomatiseringsaksen, og rundt middels når det kommer til menneskelig kontroll.

At Big Dog har en lavere score på menneskelig kontroll kan være en fordel når det kommer til denne typen robot. Dersom den kunstige intelligensen er god nok vil menneskelig påvirkning mulig redusere effekten og hensikten med teknologien når den beveger seg i et usikkert terreng. Dersom Big Dog for eksempel klarer å identifisere miner i området den beveger seg i og klarer å navigere rundt disse, vil kanskje en menneskelig kontroll ødelegge. Hadde Big Dog scoret lavere på dataautomatisering ville ressurser muligens gått tapt fordi Big Dog ikke klarte å navigere seg godt nok i terrenget den har som oppgave å komme seg trygt gjennom.

Hvordan de kommuniserer er her avgjørende, og har med hvor godt systemet forklarer valgene de gjør. Hagras (2018) definerer explainability i AI som hvordan og hvor godt et AI-basert system forklarer handlinger og prediksjoner til brukerne. Dersom Big Dog for eksempel klarer å gi beskjed til brukerne at det ligger miner i området kan resultatet bli noe annet. Altså om Big Dog klarer å forklare eller synliggjøre hvorfor den velger å gjøre som den gjør. At Big Dog har en viss grad av explainability vil jeg tro er både nødvendig og til stede da den opererer i situasjoner hvor mye står på spill.

3.1.2 Eksempel 2: Paro

Paro er en robot utformet som en sel-baby som skal hjelpe mot ensomhet og depresjon hos eldre (Subbaraman, 2013, i Phillips et al., 2016, s. 106). Paro støtter menneskets kognitive og emosjonelle ferdigheter, og krever at brukeren passer på den og viser omsorg for at den skal ha det bra. Den tilpasser seg omgivelsene og reagerer på ulike input fra brukeren. I følge Upham Ellis et al. (2005) er utformingen av denne roboten er viktig; den er søt, myk og mennesker bruker utseende for å tilegne den egenskaper selv om dette ikke er slik dyret ville oppført seg i naturen (Phillips et al., 2016).

Paro er en svært autonom robot. Ser vi den opp mot Sheridan og Verplank sine nivåer av autonomi, så er ligger den kanskje så høyt som 9. Dette nivået tilsvarer at datamaskinen informerer mennesket bare dersom den bestemmer seg for det selv. Som nevnt reagerer den utelukkende på input fra omgivelsene, og agerer deretter. Om vi ser dette i lys av Shneiderman (2020) sin todimensjonale modell, så kan man si at Paro, scorer høyt på datakontroll, og lavt på menneskelig kontroll.

Ser vi på fordelene og ulempene ved å endre grad av autonomien så vil det ha en effekt på hvordan Paro oppleves ovenfor brukeren. Når Paro har høy grad av datakontroll og lav grad av menneskelig kontroll så gjør det at Paro er mer selvstendig enn om den hadde lavere grad av datakontroll og høyere grad av menneskelig kontroll. Når hensikten med roboten er at brukeren skal ta vare på den og gi den input basert på dens egne behov, så ville den kanskje ikke hatt samme effekten ovenfor brukeren dersom man for eksempel bare kunne «skrudd av» at den var sulten. Man mister da et aspekt av innlevelse og følelsen av at babyselen faktisk trenger omsorg fra brukeren for å ha det bra. Illusjonen brytes i det den ikke har behov for deg lengre, og derfor kan det være viktig at den menneskelige kontrollen er lav.

Noe av det samme kan man si om behovet for forklaringer (explainability). Man ønsker å ta vare på selen ved å gi den det den trenger. Selvsagt trenger man et referansepunkt for hva den

mulig kan ha behov for, og hvilke funksjoner den støtter, men om den har en lampe som lyser at den er sulten fordi det er 6 timer sist den fikk mat, så mister man muligens illusjonen om at det er et ekte vesen med ekte behov. I kontrast med Big Dog, så er ikke Paro sine rolle like kritisk i sin kontekst. En bruker vil i større grad tolerere at Paro ikke er like explainable som et verktøy i militæret. Jeg vil kanskje argumentere for at det kanskje er til Paros fordel at den ikke er like forklarlig. Dette danner er lag av mystikk og kompleksitet som kanskje kan føre til at brukeren får en følelse av den er mer «ekte».

Litteraturliste

Amazon (u. å.). What is Artificial Intelligence? Hentet fra https://aws.amazon.com/machine-learning/what-is-ai/

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., ... & Teevan, J. (2019). Guidelines for human-AI interaction. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (paper no. 3). ACM. (https://www.microsoft.com/enus/research/uploads/prod/2019/01/Guidelines-for-Human-AI-Interaction-camera-ready.pdf)

Bratteteig, T., & Verne, G. (2018). Does AI make PD obsolete?: Exploring challenges from artificial intelligence to participatory design. Proceedings of the 15th Participatory Design Conference on Short Papers, Situated Actions, Workshops and Tutorial - PDC '18, 1–5. https://doi.org/10.1145/3210604.3210646

Britannica. (2020). Artificial Intelligence. https://www.britannica.com/technology/artificial-intelligence

de Caprona, Y. (2013). Empiri. I: Y. de Caprona (red.). Norsk etymologisk ordbok: Tematisk ordnet (s. 1215). Oslo: Kagge forlag.

Følstad, A. (2020). Interacting-with-ai-2020—Module-2—Session-2— handouts.pdf [Powerpoint slides]. Hentet fra: https://www.uio.no/studier/emner/matnat/ifi/IN5480/h20/Undervisningsmateriale/interacting-with-ai-2020---module-2---session-2---handouts.pdf

Følstad, A., & Brandtzæg, P. B. (2017). Chatbots and the new world of HCI. interactions, 24(4), 38-42. (https://dl.acm.org/citation.cfm?id=3085558)

Grudin, J. (2009). AI and HCI: Two Fields Divided by a Common Focus. AI magazine, 30 (4), p. 48-57

Hagras, H., Toward Human-Understandable, Explainable AI, Computer, 51, 9, 2018, p. 28-36. https://ieeexplore.ieee.org/document/8481251

Kocielnik, R., Amershi, S., & Bennett, P. N. (2019). Will You Accept an Imperfect AI?: Exploring Designs for Adjusting End-user Expectations of AI Systems. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (paper no. 411). ACM. (https://www.microsoft.com/enus/research/uploads/prod/2019/01/chi19_kocielnik_et_al.pdf)

Marr, B. (2018, 14. februar). The Key Definitions of Artificial Intelligence (AI) That Explain Its Importance. Forbes. Hentet fra https://www.forbes.com/sites/bernardmarr/2018/02/14/the-key-definitions-of-artificial-intelligence-ai-that-explain-its-importance/

Phillips, E., Ososky, S., Swigert, B. and Jentsch, F. Human-animal teams as an analog for future human-robot teams, Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Vol 56, Issue 1, (2016) pp. 1553 – 1557. DOI: https://doi.org/10.1177/1071181312561309

Shneiderman, B. (2020), Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy. *International Journal of Human-Computer Interaction*, 36:6, 495-504. DOI: 10.1080/10447318.2020.1741118

Tech Insider. (2018, 1. juli). *We Interviewed Pepper – The Humanoid Robot* [Videoklipp]. Hentet fra https://www.youtube.com/watch?v=zJHyaD1psMc

The Universal Design Project (2020). What is Universal Design? Hentet fra https://universaldesign.org/definition

Thrun, S. (2004). Toward a Framework for Human-Robot Interaction. *Human–Computer Interaction*, 19, p. 9-24.

Yang, Q., Steinfeld, A., Rosé, C., & Zimmerman, J. (2020). Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design. In Proceedings of the 2020 CHI conference on human factors in computing systems (Paper no. 164). (https://dl.acm.org/doi/abs/10.1145/3313831.3376301)

Appendix

Tilbakemelding iterasjon 1:

> Først vil jeg si at du har skrevet en grundig oppgave, med fine referanser og et godt språk. Det er veldig fin og oversiktlig struktur, og hele teksten har fin helhetlig flyt (det virker ikke som masse separate oppgaver). Du kommenterer definisjonene på en god måte, og bruker dem igjen senere i oppgaven. Veldig bra, da blir de ikke bare "hengende i løse luften".
>
> Jeg liker godt hvordan du sammenligner definisjonene av roboter ved å snakke om autonomi. Jeg har tolket autonomi som at noe kan handle på egen hånd, og ikke via et menneske for eksempel. Må det også bety at atferden ikke er bestemt (programmert) på forhånd? Din definisjon er god, du tar opp nettopp det med å handle "egenhendig".
>
> Hvis noe skal forbedres er det kanskje å drøfte litt mer rundt om AI fører bedre universelt utformede løsninger. Vil ikke et høreapparat være en spesialløsning? (Jeg syns selv dette var vanskelig å diskutere)
>
> I tillegg er det fint det du skriver om Schneidermans gyldne regler sammenlignet med Microsofts Guidelines. Kunne du diskutert dette enda mer? Som du nevner tidligere om AI får man jo systemer som lærer over tid, hvordan forholder de ulike retningslinjene seg til bruk over tid?
>
> Alt i alt, veldig bra jobba! Si fra om noe er uklart :)

Med tanke på denne tilbakemeldingen har jeg gjort følgende fra forrige iterasjon:

- Drøftet litt mer rundt AI som universelt utformede løsninger
- Utbrodert mer om sammenlikningen mellom Schneidermans gyldne regler og Microsoft sine guidelines.

Tilbakemelding iterasjon 2:

> Her er min tilbakemelding på innleveringen i modul 2:
>
> Star: Du er kjempeflink til å referere til pensum og konsepter på en naturlig måte og forklarer godt med egne ord, topp! Dette gjør at du viser god forståelse :)
>
> Star: Du svarer veldig godt i oppgave 3 om chatbots/conversational user interfaces ved å ta opp tråden på det du har skrevet om tidligere. Dette gjør at besvarelsen har en rød tråd, og det blir lettere å følge strukturen for leseren.
>
> Ønske: Du har skrevet godt om Youtube som eksempel på AI-basert system med de fire hovedkarakteristikkene, i oppgave 2.1. En ting som kunne vært forbedret var å nyansere litt hva det har å si for brukeropplevelsen. Det er kanskje ikke bare negative konsekvenser for brukeren at innholdet blir spesialtilpasset på denne måten?
>
> Igjen, veldig bra jobba, Silje! Måtte tenke meg godt om for å finne noe som kunne forbedres ...

Med tanke på denne tilbakemeldingen har jeg gjort følgende fra forrige iterasjon:

- Skrevet mer om hvorfor algoritmene er tilpasset som de er for å nyansere bildet mer.