

Interaksjon med AI

Del 1 - Definisjoner/teori

1. Tre definisjoner: AI

- 1) Liseter (2018) definerer kunstig intelligens som et “forsknings- og utviklingsfelt innenfor datateknologien som benytter teoretiske og eksperimentelle dataverktøy til å studere intelligent atferd, og som bruker resultatene til å konstruere datasystemer som er «intelligente» i den forstand at de er i stand til å løse problemer og lære av egne erfaringer.”
- 2) I følge Rowe (2018) kan man forstå kunstig intelligens som utviklingen av systemer som “er i stand til å gjennomføre oppgaver som vanligvis krever menneskelig intelligens.”
- 3) Cambridge Dictionary (udatert) sier at kunstig intelligens er “computer technology that allows something to be done in a way that is similar to the way a human would do it.” Parnas (2017) hevder på en annen side at begrepet “kunstig intelligens” er et *buzzword* som mange tror de kan, men som ingen klarer å definere.

Jeg synes disse definisjonene på hver sin måte er noe vage, samtidig som de kommer inne på flere av aspektene jeg selv synes er viktig å poengtere når man skal definere AI, da hovedsakelig problemløsning og forståelse av menneskelig intelligens. Slik vi har diskutert begrepet AI i kurset, har fokuset vært på å forstå AI-systemer som systemer som imiterer menneskelig oppførsel. At oppførselen er menneskelig-lik, vil si at man bygger systemet på mange av de samme prinsippene som menneskelig kognisjon. Definisjonene over nevner ingenting om hvor intelligent et system kan være, men i kurset skiller vi mellom *super*, *general* og *narrow* intelligens, hvor interaksjon med AI typisk skjer ved sistnevnte.

1.2. Når ble begrepet “AI” først brukt?

Minsky (1986) hevder den moderne forskningen av kunstig intelligens på 1950-tallet, hvor Smith, McGuire, Huang & Yang (2006) mener John McCarthy brukte begrepet første gang i 1956.

2. Tre definisjoner: Robotikk

- 1) I følge Hill (2017) er robotikk en gren av teknologi som omhandler roboter. Videre hevder han at roboter er “programmable machines which are usually able to carry out a series of actions autonomously, or semi-autonomously.”

- 2) Montebelli, Billing, Lindblom & Dahlberg (2017) diskuterer robotikk i lys av konseptet robot, da omtalt som “an artifact whose configuration of sensors, actuators, and integrated control system provides a significant level of flexible, independent, and autonomous action.”
- 3) Siegel (2015) presiserer på lik linje som Montebelli et al. (2017) konseptet robot i lys av robotikk, og definerer en robot som “a machine that senses, thinks, and acts.”

3. Tre definisjoner: Maskinlæring

- 1) Parnas (2017) snakker om maskinlæring som en tilnærming til kunstig intelligens hvor man bygger programmer som i utgangspunktet har minimal kapasitet, men som forbedrer utførelsen sin etterhvert som de blir brukt.
- 2) Domingos (2012) definerer maskinlæring som systemer som “automatically learn programs from data.”
- 3) Marr (2016) hevder at maskinlæring er “a current application of AI based around the idea that we should really just be able to give machines access to data and let them learn for themselves.”

Slik jeg ser det, vil definisjonen til både Parnas (2017) og Domingos (2012) inneholde mange av de samme elementene som benyttes når vi diskuterer begrepet “maskinlæring” i kurset. Her snakker vi om maskinlæring som en metode innenfor kunstig intelligens, hvor det er viktig med store mengder data og mye datakraft for at utførelsen til maskinlæringsmodellene skal være best mulig. På denne måten kan modellene forbedres under bruk, med andre ord lære fra dataen, og videre bli “mer intelligente”.

4. Forholdet mellom AI og robotikk

Slik jeg forstår det, er robotikk og kunstig intelligens to ulike grener av teknologi. For meg handler robotikk om å designe og konstruere roboter. Robotene er programmerbare objekter og kan vanligvis utføre handlinger autonomt eller semi-autonomt. Forholdet mellom kunstig intelligens og robotikk blir derfor synlig ved kunstig intelligente roboter, der roboten ikke lenger er kontrollert av mennesker, men i større grad styrer seg selv. På denne måten kan roboten utføre oppgaver på lik linje som mennesker, sammen med mennesker.

5. Min forståelse av AI

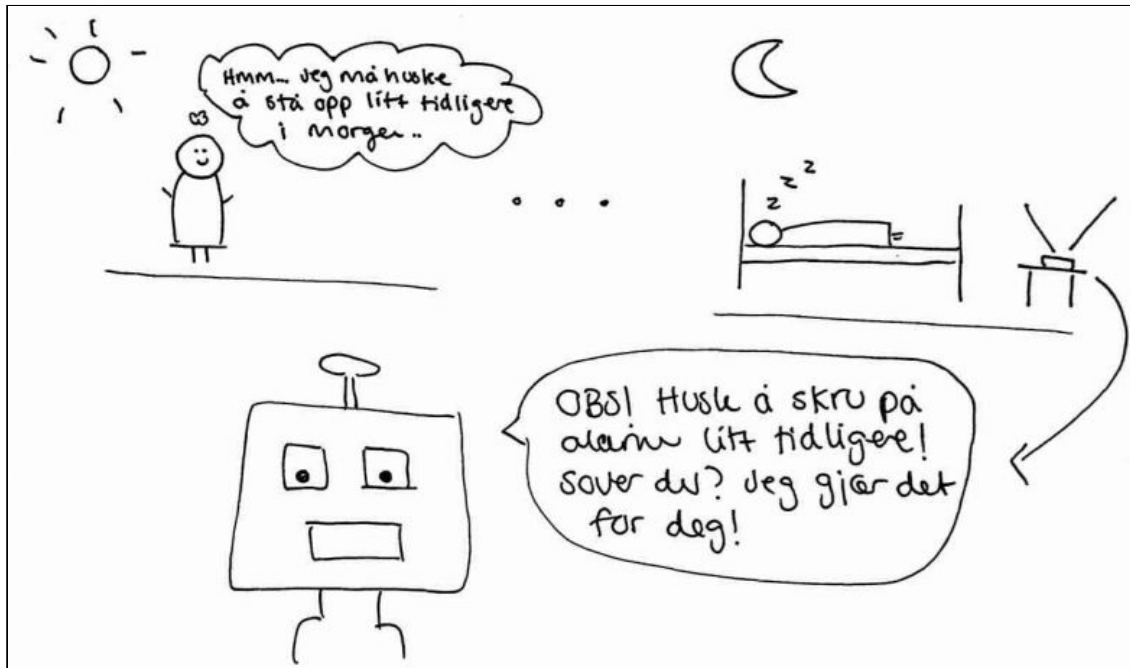
For meg handler kunstig intelligens om å forstå menneskelig kognisjon, eksempelvis tenkning, resonnering, beslutningstaking, persepsjon og bevegelse, og hvordan man utnytter denne kunnskapen til å utforme systemer med de samme egenskapene. Kunstig intelligens er med andre ord et bredt konsept der ønsket er å utvikle maskiner som kan utføre oppgaver på en måte som vi vil betrakte som "smarte". Slik jeg ser det vil både maskinlæring og robotikk være ulike grener av kunstig intelligens, men det må ikke være det.

I lys av min forståelse av maskinlæring, vil kunstig intelligens være det vi ønsker å oppnå, og maskinlæring en metode man kan bruke for å oppnå det målet. Det er to forutsetninger for maskinlæring; store datasett og stor datakraft. En enkel forståelse av kunstig intelligens med bruk av maskinlæring kan derfor være at det er systemer som lærer og forbedrer seg på basis av store sett med data.

11. Min forståelse av autonomi

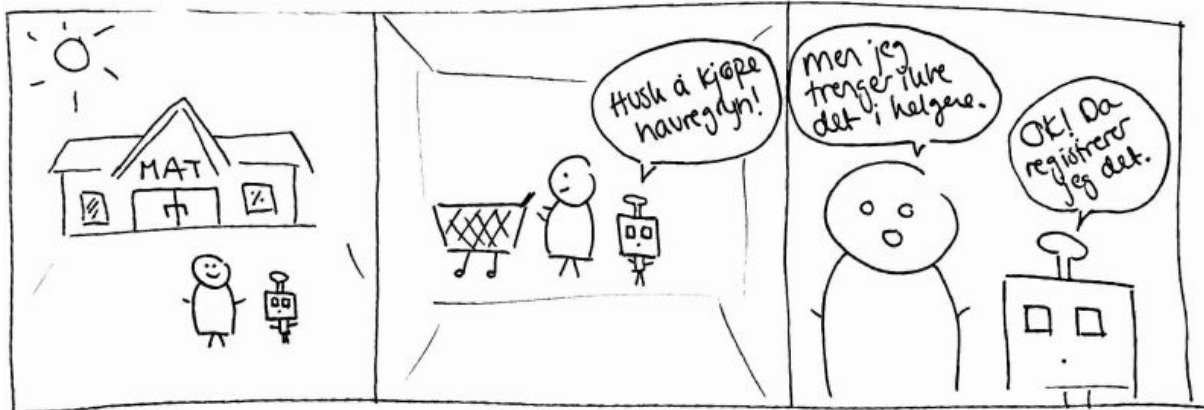
Å være autonom forstår jeg som muligheten til å bestemme over seg selv og å være selvstendig, med andre ord å kunne ta valg uavhengig av andres påvirkning. Basert på denne forståelsen mener jeg at autonomi for mennesker og maskiner vil bety mye av det samme, ettersom det handler om at både mennesket og maskinen kan fatte beslutninger av seg selv. Jeg tenker imidlertid at *intensjonen* bak beslutningen eller begrunnelsen for hvorfor en handling blir gjort, er viktige aspekter som kan skille de to. Maskiner tar beslutninger basert på data og regler, mens menneskelig beslutningstaking er mer kompleks og påvirkes også av eksempelvis erfaring og opplevelse. Jeg opplever derfor at hvorvidt et system er autonomt avhenger av menneskelig involvering, mens for et menneske vil autonomi være utgangspunktet.

6. Tegning



Tegningen skal etterligne et system som minner meg på alle ting jeg tenker jeg skal gjøre i løpet av dagen, men som jeg har glemt. Systemet har lært seg vanene og rutine mine, og kan eksempelvis be meg sette på en alarm dersom jeg har glemt det, eller minne meg på å kjøpe varer jeg trenger når jeg er på butikken.

Ved utvikling av interaksjonsbaserte AI-systemer, er det spesielt tre nøkkelord som er viktige: læring, forbedring og store datasett. Med *læring* mener vi at systemene må utformes dynamiske slik at de enkelt kan endres. Et designprinsipp vil derfor være at systemet alltid må vise overfor brukeren hva som er mulig og hva som er begrenset, nettopp fordi dette ofte kan endre seg. Dette er også viktig for å sette riktige forventninger hos brukeren. Med *forbedring* menes at systemene må utformes slik at feil vil skje. Et system som gjør feil, er et system som forbedres. Dette kan for eksempel gjøres ved å designe for usikkerhet (eksempelvis vise overfor brukeren at systemet er usikker på resultatet), eller la brukeren gi konkrete tilbakemeldinger på hva som er feil ("Var denne informasjonen nyttig?"). For å muliggjøre disse designprinsippene, er det derfor viktig å utforme systemet slik at det får inn *store datasett*. Dataen må med andre ord hentes gjennom brukers interaksjon med systemet, gjerne ved å vise overfor brukeren at vedkommende har nytte av å gi mer data.



Tegningen viser et grensesnitt for et tenkt AI-system som foreslår hvilke matvarer jeg trenger på butikken, basert på hva som er mine mest kjøpte matvarer og sannsynligheten for at jeg trenger å kjøpe noe på nytt. Eksempelen under illustrerer at systemet lærer om varer jeg ikke trenger gjennom min tilbakemelding på at dette er feil.

Del 2 - Artikler

7 & 8. On the Subject of Objects

I artikkelen *On the Subject of Objects* presenterer Susi & Ziemke (2005) fire perspektiver på hvordan vi forstår teknologi, og forholdet mellom menneske og maskin. Forfatterne har fokus på relasjonen mellom en agent og dets miljø, og spesielt hvordan subjekter oppfatter objekter eller artefakter og hvordan disse (kan) brukes. Fire ulike perspektiver blir presentert; *functional tone*, *equipment*, *affordance* og *entry point* og kommer fortrinnsvis fra biologi, filosofi, psykologi og kognisjonsvitenskap. Forfatterne hevder at til tross for at disse har mange likheter, har de også flere forskjeller. Dette mener de er viktig å være klar over for å

forstå menneskelig kognisjon og interaksjon med teknologi, samt hvordan roboter interagerer med miljøet rundt seg.

Det første perspektivet som blir presentert, *functional tone*, tar utgangspunkt i biologi og Jakob von Uexküll. Uexküll hevder at ethvert dyr tilskriver mening til de fysiske objektene det møter på, og på den måten skaper sitt eget individuelle univers ved å “forme” verden til seg selv. The subjektive universet kaller han for *Umwelt*, og er en lukket enhet som kun består av subjektets perseptuelle verden. Når subjektet da skal interagere med et objekt, samt forstå hvordan interaksjonen skal foregå, må objektet ha det Uexküll kaller *functional tone*.

Slik jeg forstår det, handler *functional tone* om hvordan vi opplever et objekt eller en artefakt, samt hvilken informasjon vi tildeler det objektet som gjør at vi forstår hvordan vi skal interagere med det. Ifølge Uexküll er objektene alltid nøytrale, men får tildelt en “betydning” eller perseptuelle hint når de er i interaksjon med subjektet, av subjektet. Subjektet skaper da en mening om hvordan de kan bruke objektet, og objektet får da en *functional tone*. Hva som er *functional tone*, avhenger av subjektet og for eksempel humøret til subjektet. Eksempelvis kan en stol brukes til å sitte på, men også som et våpen, avhengig av hvilken betydning subjektet tildeler objektet.

Det andre perspektivet som blir presentert er *equipment*, og kommer fra filosofen Martin Heidegger. Til forskjell fra Uexküll fokuserer Heidegger mer på mennesker, samt det å eksistere. Han hevder det er en gjensidig avhengighet mellom subjektet og objektet, og at disse ikke kan bli sett på som to isolerte enheter. Subjektets forståelse av hvordan man skal bruke et objekt er med andre ord avhengig av subjektets egne handlinger, og hvordan vedkommende kan manipulere objektet, ettersom manipulasjonen vil si noe om hvordan objektet kan brukes. Objektene må videre bli definert ut fra en kontekst, siden det er kun innenfor en meningsfull kontekst at objektet oppfattes som det gjør.

Affordance er det tredje perspektivet som blir presentert. Slik jeg forstår *affordance* så handler dette om hvordan objektets attributter gir subjektet hint om hvordan det kan brukes. Konseptet kommer fra psykologen James J. Gibson og refererer til relasjonen mellom fysiske objekter og et subjekt. Denne forståelsen av interaksjon mellom menneske og teknologi er mye lik Uexküll, men til forskjell fra Uexküll som hevder betydningen til objektet endres ut fra humøret eller motivasjonen til subjektet, mener Gibson at *affordance* er objektivt og aldri endres. Dette kan med andre ord tilsi at *affordance* noen ganger blir mistolket, spesielt ettersom bruk av objekter blir formet av sosiale normer, kunnskap og erfaringer.

Det fjerde perspektivet som blir presentert er *entry points* (Kirsh, 2001). *Entry points* forstår jeg på lik måte som *affordance*, da de inviterer subjektet til å gjøre noe. Eksempelen som blir brukt er hvordan man strukturerer kontorpulten sin, og hvordan ulike elementer eller objekter

på pulten har ulike karakteristikk som gjør at man i ulik grad benytter seg av entry points for å for eksempel gjøre oppgaver på en bedre måte. Slik jeg ser det vil *entry points* med andre ord si noe om hvordan miljøet vi er i påvirker hvordan vi tenker på, og det oppstår da en gjensidighet mellom subjektet og objektet (*entry points*).

9. Humans and Automation: Use, Misuse, Disuse, Abuse

I denne artikkelen presenterer Parasuraman & Riley (1997) både teoretiske, empiriske og analytiske studier av hvilke faktorer som påvirker menneskelig bruk av automatisering. Forfatterne definerer automatisering som maskinbasert utførelse av en handling som tidligere ble gjort av et menneske (s. 231), og diskuterer både “use, misuse, disuse, abuse” av automatiseringsteknologi. *Use* refererer her til den faktiske bruken, eller unnlattelse av bruk, av automatisering. Hvorvidt man velger å bruke automatisering eller ikke, mener forfatterne påvirkes av eksempelvis tillit, risiko og arbeidsmengde. *Misuse* handler om å ha for stor tillit til automatisering. Dette kan eksempelvis være å benytte automatisering der det ikke passer, noe som kan resultere i bias knyttet til beslutninger som blir tatt av systemet. *Disuse* handler om forsømmelse, for eksempel å slå av automatiserte alarmer eller sikkerhetssystemer på grunn av mistillit, og *abuse* referer til upassende automatisering av funksjoner. Innføringen av denne typen automatisering kan for eksempel være på grunnlag av økonomisk gevinst, da uten å nødvendigvis tenke på konsekvensene det har for menneskelig ytelse eller bruk. Dette kan videre lede til både *misuse* og *disuse*. Forfatterne presenterer videre ulike tiltak for å unngå ulik misbruk av automatisering, hvor åpenhet og bevisstgjøring rundt de ulike formene for misbruk er gjennomgående.

10. Film: WestWorld

WestWorld er en serie fra HBO som er basert på en film med samme navn fra 1973. Handlingen foregår i en fornøylespark for voksne med den ville vesten som tema, hvor attraksjonene er kunstig intelligente roboter. De besøkende kan leve ut sine fantasier og gjøre hva enn de ønsker med robotene, uten at det får konsekvenser. Rent utseendemessig er robotene like mennesker, og interaksjonen er også veldig lik menneskelig interaksjon. Denne interaksjonen utvikler seg underveis i serien, da robotene etterhvert tilegner seg egenskaper som er mer og mer lik menneskelig intelligens. Etterhvert er det vanskelig (både som seer og besøkende) å skille robotene og menneskene fra hverandre, og relasjonene og maktforholdet dem i mellom blir også snudd på hodet.

13 & 14. Spørsmål

Spørsmål til artikkelen *What we talk about when we talk about context* (Dourish, 2004):

- Hvilke implikasjoner har denne forståelsen av kontekst på forståelse av bruk av teknologi som ikke nødvendigvis er *ubiquitous computing*?

Spørsmål til artikkelen *Context-awareness* (Shilt, Hilbert & Trevor, 2002):

- Hvordan skal man ta hensyn til personvern når man hele tiden bruker informasjon om en persons lokasjon, miljø og sosial situasjon for å fasilitere kommunikasjon? Hvor går grensen mellom personvern og gode brukeropplevelser?

15. Like Having a Really Bad PA

I denne artikkelen undersøker og presenterer Luger & Sellen (2016) faktorer som motiverer og begrenser bruk av konverserende agenter (CA). De mener det er store forskjeller mellom brukernes forventninger til et slikt system og hva systemet faktisk kan gjøre, og hevder det dermed er viktig å få en forståelse av “everyday use” og hvilke faktorer som påvirker god og dårlig bruk. Disse faktorene ser de i lys av Norman sin “gulf of execution and evaluation”; forstått som i hvilken grad systemet kan oppfattes og tolkes av brukeren, så vedkommende får nøyaktige forventninger og forstår hensikten med bruk. Jo smalere “gulf”, desto bedre brukeropplevelse.

Forfatterne presenterer flere aspekter som påvirker hvordan brukerne opplever å interagere med de konverserende agentene. Dette er spesielt knyttet til brukernes kunnskap om hva systemet faktisk kan gjøre, den faktiske intelligensen til systemet, og hva som er målet med bruk. For å møte disse utfordringene, foreslår forfatterne først og fremst å designe systemer som *viser* brukeren systemets intelligens. Dette er spesielt viktig der interaksjon skjer med brukere med mindre kjennskap til teknologi, og som dermed har høyere forventninger til hva systemet kan gjennomføre. Videre presenterer forfatterne bruk av humor i interaksjonen, samt undersøke hvordan humor kan støtte opp under brukerens forventning til systemets intelligens. Brukernes mentale modell om menneskelig interaksjon vil påvirke hvordan de interagerer med systemet, og forfatterne mener i den anledning at det også er viktig å undersøke hvordan de kan vise systemets muligheter og begrensninger gjennom *andre* instanser enn kun når systemet mislykkes i å fullføre en oppgave. Til slutt foreslår forfatterne også å fokusere mer på hva som er den faktiske bruken av systemet, slik at hvordan systemet interagerer, utfører oppgaver og gir informasjon til brukeren understøtter den faktiske bruken.

Slik jeg ser det er mange av disse forslagene som vil være gjeldene for utforming av AI-baserte systemer generelt. Brukere vil alltid bringe inn sine mentale modeller i interaksjonen med et system, og hvor utviklet disse modellene er avhenger av brukernes kunnskap og kjennskap til teknologi. Dette resulterer i hvorvidt brukeren opplever tillit til systemet, hvor mangelfull kunnskap for eksempel resulterer i at flertallet ikke stoler på at systemet kan gjennomføre sensitive oppgaver. Jeg tenker derfor at spesielt forslag 1 (å vise systemets intelligens og 4 (å vise systemets kapasitet) er hensiktsmessig uavhengig av om det snakk om utvikling av en konverserende agent eller om det er utformingen av en robot.

16. Using Artificial Intelligence to Augment Human Intelligence

Carter & Nielsen (2017) diskuterer i denne artikkelen ulike syn på systemer og maskiner, og presenterer et nytt felt kalt *artificial intelligence augmentation* (AIA), som en syntese av kunstig intelligens (AI) og *intelligence augmentation* (IA). Forfatterne mener at selv om det alltid har vært overlapp mellom AI og IA, har IA typisk fokusert på å bygge systemer som setter mennesker og maskiner sammen, mens AI har fokusert på fullstendig outsourcing av intellektuelle oppgaver til maskiner. AIA vil med andre ord være en kombinasjon, hvor man bruker AI-systemer for å utvikle nye metoder for “intelligensforbedring” hos mennesker.

Tradisjonelt har datamaskiner først og fremst blitt sett på som verktøy for å løse numeriske problemer. Etterhvert som teknologien har utviklet seg, mener forfatterne at det nå er velkjent at AI vil forandre hvordan vi samhandler med datamaskiner. Forfatterne presenterer derfor et annet syn, hvor AI faktisk endrer menneskeheten, og hvor AI hjelper oss med å utforme nye kognitive teknologier, som igjen utvider spekteret av menneskelig tenkning. Forfatterne presenterer i den anledning bruk av *generative grensesnitt* som en metode for å forbedre intelligens; grensesnitt som kan brukes til å utforske og visualisere generative maskinlæringsmodeller. Ved å lage slike brukergrensesnitt som bruker representasjonene som finnes i maskinlæringsmodellene, argumenterer forfatterne at de kan forbedre menneskelig intelligens, herunder gi mennesker nye verktøy for å resonnerer og være kreative.

17(16). Levels of automation

Jeg forstår automatisering som en prosess med formål å gjøre en handling automatisk, slik at handlingen i større eller mindre grad styres av seg selv. Slik jeg forstår det, kan det være hensiktsmessig å dele automatisering inn i: ingen grad (mennesket utfører alt), liten grad (systemet utfører noen handlinger, da hovedsakelig for å støtte opp under menneskelig arbeid), stor grad (systemet utfører de fleste handlinger, og mennesker hjelper til ved unntak) og høy grad (systemet utfører alt). Slike nivåer blir også introdusert hos Sheridan & Verplank (1978).

Fordeler med høy grad av automatisering vil være effektivitet. Maskinbasert utførelse er hurtigere enn hva mennesker er skapt til å klare, kan derfor øke kapasiteten på handlingen som blir gjort. Utførelsen blir videre mer nøyaktig ettersom den baserer seg på de samme variablene hver gang, og er derfor mer objektivt og pålitelig enn manuelt arbeid. Dette tilsier også at systemet kan gjenta samme prosess mange ganger uten å gjøre feil, noe som kan være utfordrende for et menneske. Dette er eksempelvis en stor fordel i kritiske systemer hvor konsekvensene av feilhandlinger er store. Automatiserte systemer kan videre anses som mer stabile enn mennesker, ettersom systemene ikke blir syke eller drar på ferie. Nedetid blir derfor også redusert.

Ulemper ved automatisering er først og fremst håndtering av avvik. Etersom systemet ofte bare kan utføre oppgaver det er satt til å gjennomføre, kan det derfor ikke håndtere unntakstilfeller uten menneskelig hjelp. Dersom mange arbeidsprosesser er basert på automatiserte prosesser, kan det videre oppstå store utfordringer dersom noen av prosessene svikter. Ved bruk av automatisering i offentlig sektor kan det være vanskelig for borgere å forstå oppgaver de må gjøre, eller prosesser som blir gjennomført, for eksempel ved skatt. Automatisering vil videre erstatte menneskelig arbeidskraft, noe som kan være negativt for eksempel i helse- og omsorgssektoren. Det er videre potensielt mye kostnader tilknyttet automatisering, ettersom det kan være tidkrevende å utforme gode og robuste prosesser.

Referanser

Cambridge Dictionary. (udatert). AI. Hentet fra:

<https://dictionary.cambridge.org/dictionary/english/ai>.

Carter, S., & Nielsen, M. (2017). Using artificial intelligence to augment human intelligence. *Distill*, 2(12). <https://doi.org/10.23915/distill.00009>.

Domingos, P. (2012). A Few Useful Things to Know about Machine Learning. *Comm. of the ACM*, 55(10), 78-87.

Dourish, P. (2004). What we talk about when we talk about context. *Personal and ubiquitous computing*, 8(1), 19–30.

Hill, O. (2017, 19. juli). What's the Difference Between Robotics and Artificial Intelligence? Hentet fra:

<https://blog.robotiq.com/whats-the-difference-between-robotics-and-artificial-intelligence>.

Liseter, I. M. (2018, 20. februar). Kunstig intelligens. Hentet fra:

https://snl.no/kunstig_intelligens.

Luger, E., & Sellen, A. (2016). Like having a really bad PA: the gulf between user expectation and experience of conversational agents. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 5286-5297.

Marr, B. (2016, 6. desember). What Is The Difference Between Artificial Intelligence And Machine Learning? Hentet fra:

<https://www.forbes.com/sites/bernardmarr/2016/12/06/what-is-the-difference-between-artificial-intelligence-and-machine-learning/#3a5ef7fd2742>.

Minsky, M. (1986). *The Society of Mind*. New York: Simon & Schuster.

Montebelli, A., Billing, E., Lindblom, J. & Dahlberg, G. M. (2017). Reframing HRI Education: A Dialogic Reformulation of HRI Education to Promote Diverse Thinking and Scientific Progress. *Journal of Human-Robot Interaction*, 6(2), 3-26.

Parasuraman, R. & Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors: The Journal of the Human Factor and Ergonomics Society*, 39, 230–253. <https://doi.org/10.1518/001872097778543886>.

Parmas, D. L. (2017). The Real Risks of Artificial Intelligence. *Communications of the ACM*, 60(10), 27-31.

Rowe, J. (2018). Tre ting du må vite om kunstig intelligens (AI). Hentet fra:

<https://www2.deloitte.com/no/no/pages/technology/articles/tre-ting-vite-kunstig-intelligens-ai.html>.

Schilit, B. N., Hilbert, D. M. & Trevor, J. (2002). Context-aware communication. *IEEE Wireless Communications*, 9(5), 46–54.

Sheridan, T. B. & Verplank, W. (1978). *Human and Computer Control of Undersea Teleoperators*. Cambridge, MA: MIT Press.

- Siegel, M. (2015, 3. juli). What is the definition of a robot? Hentet fra:
<http://serious-science.org/what-is-the-definition-of-a-robot-3587>.
- Smith, C., McGuire, B., Huang, T. & Yang, G. (2006). *The History of Artificial Intelligence*. University Of Washington. Hentet fra:
<https://courses.cs.washington.edu/courses/csep590/06au/projects/history-ai.pdf>.
- Susi, T. & Ziemke, T. (2005). On the Subjects of Objects: Four Views on Object Perception and Tool Use, *Triple C Journal for a Global Sustainable Information Society*, 3(2), 6-19. <https://doi.org/10.31269/triplec.v3i2.19>.