

INDIVIDUELL OPPGAVE - IN548

Andre iterasjon – Emil Säll Fuglerud (ehfugler)

Oppgave 1. Måter å definere kunstig intelligens

Selv om det er en hype rundt kunstig intelligens i dag (Artificial Intelligence), har forskning på området foregått i over 50 år. Alan Turing var en av de første til å utforske konseptet om intelligente maskiner når han skrev *Computing Machinery and Intelligence* (Turing, 1950). Turing definerte hva mange fortsatt anser som det ultimate målet for å oppnå kunstig intelligens, der en intelligent maskin kan defineres ved sin ”ability to achieve human-level performance in all cognitive tasks, sufficient to fool an interrogator” (Russel & Norvig, 1995, p. 5).

En annen måte å definere kunstig intelligens er ved å referere til spesifikke teknologier. I dag er maskinlæring og *deep learning* de mest brukte teknologiske tilnærmingene. (Smith et al., 2006). Kunstig intelligens kan derfor eksempelvis defineres til som en maskin som gjennom å identifisere mønstre eller regler i fortiden (data) kan bruke dem til å gjøre gjetninger i fremtiden (data), noe som gir den stor prediktiv kraft (Hutter, 2007, p. 241).

Ved bruk av denne typen teknologi, demonstrerer maskiner noe som kan kalles for *narrow intelligence*, der maskinen defineres som intelligente innenfor spesifikke domener – eksempelvis gjennom å identifisere bestemte objekter i bilder, kunne kjenne igjen og forstå språk, spille sjakk osv. Dette kan skilles fra *strong intelligence* eller *artificial general intelligence (AIG)* der maskiner demonstrerer en høyere nivå av intelligent adferd ved at de “possess a reasonable degree of self-understanding and autonomous self-control, and have the ability to solve a variety of complex problems in a variety of contexts, and to learn to solve new problems that they didn’t know about at the time of their creation.” (Goertzel & Pennachin, 2007, p. 1).

Nytt i iterasjon 2:

I forelesningen ble det presentert en måte å skille mellom ulike nivåer av kunstig intelligens. Dette var noe jeg også gjorde i obligen, men på forelesning ble et tredje nivå introdusert; *Artificial super intelligence*. Slik jeg forstod det, vil kunstig intelligens på dette nivået overgå menneskers intelligens, og at det derfor også vil være vanskelig å vite hvilken form for intelligens dette vil være.

I forelesning ble det også referert til spesifikke teknologier, som her snakket om som ulike tilnærminger til å oppnå kunstig intelligens. Dette ser også ut til å legge grunnlaget for én måte å forstå kunstig intelligens i dag; «a computer system learning and improving on the basis of large datasets». Med fremtidige teknologier som representerer helt andre måter å oppnå en form for kunstig intelligens vil vi kanskje få helt andre definisjoner.

Oppgave 2. Måter å definere robotikk

Et søk på Wikipedia (2015) viser at robotikk kan defineres som et fagfelt eller studie som omhandler design, konstruksjon og bruk av roboter. I litteraturen om robotikk finner vi mer

utfyllende definisjoner, der Siciliano et al (2010) definerer robotikk som studiet av "those machines that can replace human beings in the execution of a task, as regards to both physical activity and decision making." Denne definisjonen viser også til at robotikk er et studie eller fagflet, men ordlyden skiller seg fra wikipedia sin definisjon ved at de beskriver det som studiet av maskiner som tar over menneskers oppgaver fremfor studiet av roboter. Man kan derfor undre seg over om det er intensjonelt at de ikke eksplisitt nevner roboter.

LEO Robotics, som driver med innovasjon og utvikling av roboter mener at dette avhenger av hvordan du definerer roboter. "An autonomous, self-driving car may or may not be a robot, depending on your interpretation of the definitions, but robotics is definitely involved in its design process." De på sin side definerer robotikk som "the study of mechanical engineering, electrical engineering, electronic engineering and computer science and is a broader way of looking at developments". (Lego Robotics, 2015).

Oppgave 3. Definisjon av maskinlæring

Som tidligere nevnt er maskinlæring (ML) en av de mest brukte tilnærmingene for å oppnå intelligent adferd i maskiner, og wikipedia definerer ML nettopp som et tverrfaglig fagområde som utgjør en gren av kunstig intelligens. Slik begrepet tilsier handler det i korte trekk om maskiner som lærer, og mer spesifikt definerer Wikipedia det som "en vitenskapelig disiplin opptatt av design og utvikling av algoritmer som gjør datamaskinen i stand til å lære og utvikle adferd basert på empiriske data" (Wikipedia, 2018).

En annen tilsvarende definisjon gis av Domingo (2012) som beskriver ML som algoritmer som automatisk lærer maskinen å utføre viktige oppgaver ved å gjøre generaliseringer basert på store mengder data. For å kunne benytte seg av ML kreves det altså at man har store mengder data tilgjengelig og at man kan benytte seg av statistiske teknikker for å gjøre generaliseringer basert på dem.

Google gir på sine sider denne definisjonen på ML: "*Machine learning* is functionality that helps software perform a task without explicit programming or rules. Traditionally considered a subcategory of artificial intelligence, machine learning involves statistical techniques, such as *deep learning* (aka neural networks), that are inspired by theories about how the human brain processes information" (Google, 2017). Det finnes altså ulike typer av maskinlæring som også kan beskrives som ulike former for statistiske metoder, der *deep learning* er et eksempel på dette.

Nytt i iterasjon 2:

I kurset blir det diskutert mye av det samme som definisjonene gitt i obligen. Det handler om å gjøre generaliseringer basert på store datasett, uten å eksplisitt programmere inn regler. I kurset har fokusert vært på deep learning, som er en form for maskinlæring. Slik jeg forstod det, er det antall lag som skiller de to. I definisjonene gitt i obligen ble det ikke referert til hvilken spesifikk adferd maskinen kan lære, i forelesning ble det fokusert på klassifisering av bilde og tekst. Det ble også tydelig for meg hvilke variabler programmereren kan justere slik

som antall lag og nevroner, og at mye av logikken, eks bak den optimale vekten, kan være skjult. (black box)

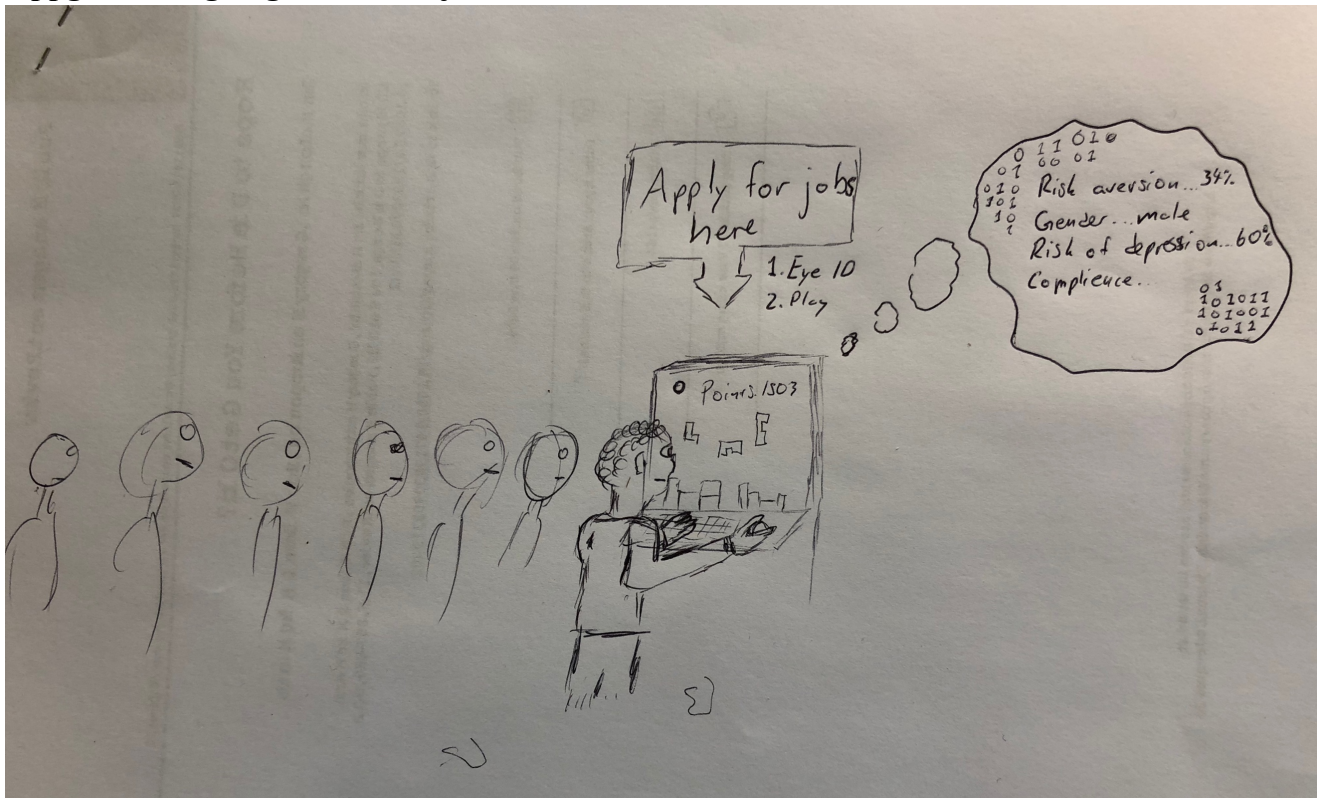
Oppgave 4. Robotikk vs. kunstig intelligens

AI robotics er et begrep som brukes om roboter som innehar en form for intelligens. For meg virker det som om at alle maskiner som innehar en form for intelligens, slik som en *chat bot* eller lignende, kan defineres som en robot, men at alle roboter kan ikke defineres som en AI.

Oppgave 5. Min egen definisjon av kunstig intelligens

I sin enkleste form kan en maskin defineres som å utøve intelligent adferd om den på egenhånd kan lære seg å løse oppgaver innenfor et bestemt domene. Maskinlæring er derfor én tilnærming til kunstig intelligens fordi det gir maskinen evnen til å lære fra fortiden (data) og på egen hånd bygge opp interne representasjoner og logikk for å utføre oppgaver, eksempelvis prediksjon eller klassifisering.

Oppgave 6. Tegning av interaksjon med AI.



Tegningen viser en illustrasjon av en dystopi der arbeidsledige i et fattig område kan gå til en AI for å bli vurdert om de er gode eller dårlige kandidater for å få jobb. Ved å spille ulike spill og biometrisk registrering blir de vurdert for alt fra intelligens til depresjon.

Nytt i iterasjon 2:

1. Design for the systems to be changing

- Explain dynamic character of the system

- Show capabilities – to make clear to the user what it can do continuously
- Clear on limitations – set the limitations of the interaction right.

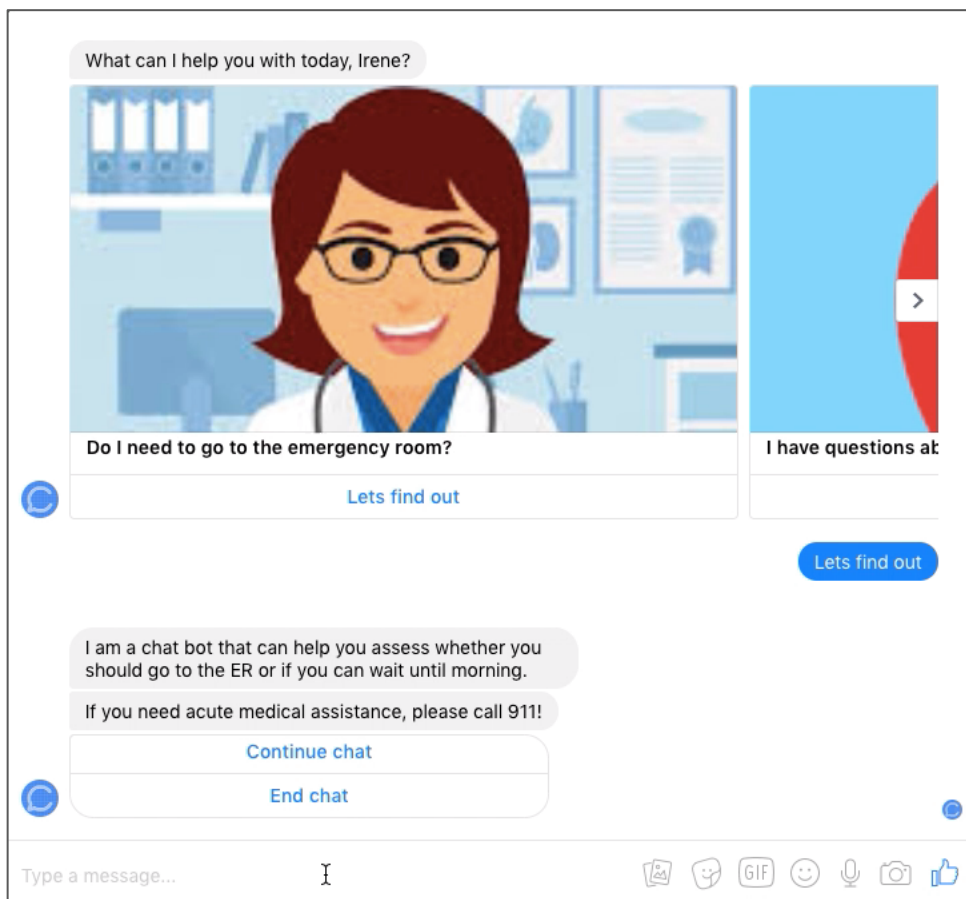
2. Design for uncertainty (Systems improve as they get more data, making them more adapted to you – making a mistake is an inherent part of improving.)

- Inform and adapt depending on certainty level
- Make recovery easy
- providing a way for learning from mistakes

3. Design for data capture (as part of the system)

- Make the users benefit from the data you are collecting
- Privacy by design

(A. Følstad, 2018).



Skissen er fra chatbot laget i kruset. Illustrere hvordan den forsøker å vise brukeren sine evner og begrensninger i samtalen.

Oppgave 7 – 8. On the Subject of Objects: Four Views on Object Perception and Tool Use

Artikkelen tar for seg og sammenligner fire ulike måter å forstå relasjonen mellom subjekt og objekt (aka. interaksjonen mellom en agent og sine omgivelser), og stiller seg kritisk til at disse ofte blir brukt om hverandre, og gjerne slått sammen under *affordance* begrepet. Ved en nærmere titt mener de at man kan identifisere en del forskjeller mellom de ulike perspektivene. Disse forskjellene blir spesielt viktig siden det å forstå relasjonen mellom subjekt og objekt er avgjørende for å forstå menneskers kognisjon (hvordan forstår vi bruken av noe) og interaksjon med teknologi. Dette gjelder også for å forstå roboters interaksjon med sine omgivelser.

Til forskjell fra *Functional Tone* (Se nedenfor) mener Heidegger at subjekt og objekt ikke kan ses som to adskilte enheter, siden de er gjensidig avhengige av hverandre. Han er også uenig i at relasjonen mellom subjekt og objekt konstitueres av noe mentalt (intensjonal, representational osv.) og mener isteden at den grunnleggende formen for å eksistere er å være i relasjon til andre ting. Equipment er det det er når det yter sin funksjon, som kommer til synet for oss først når det manipuleres og brukes. *Affordance* på sin side ser også forbi subjekt-objekt distinksjonen og handler om hva det fysiske miljøet objektivt sett tilbyr/afford et menneske/dyr i kraft av sin substans og overflate. For et menneske kan en stol sittes på, for en bil kan en stol klatres på osv. En penn vil alltid kunne brukes/afford å drepe et annet menneske, selv om en person subjektivt sett ikke ser det eller bruker det som et drapsvåpen. *Entry point* er et teoretisk konsept for å beskrive hvordan mennesker løser kognitive problemer ved å manipulere verdenen heller enn å tilpasse seg den selv. Eksempelvis for å redusere kognitiv belastning. *Entry point* brukes her for å oppnå det Kirsh beskriver som *cognitive affordance*. Eksempler på dette kan være en bunke med papirer, en gjøremålsliste osv, som gir hint og inviterer til å utføre en oppgave eller om å tre inn i et *information space*.

Functional Tone

Et konsept utviklet av Von Uexküll, som springer ut av en forståelse av kunnskap som noe subjektivt konstruert, der fysiske objekter tilegnes mening tilpasset vår subjektive virkelighet. Han benytter seg av begrepet *functional tone* til å forklare hvordan et menneske kan forstå bruken av et fysisk objekt, eksempelvis at en penn kan brukes til å skrive med. Når et individ først blir introdusert til en penn vil individet se dens fysiske attributter uten at det gir noen videre mening om bruken/funksjonaliteten, noe han kaller for "receptor image". Ved å se hvordan noen andre skriver med en penn vil han tilskrive pennen ny mening i form av nye egenskaper, som kalles *functional tone*. Når det oppstår en slik relasjon mellom subjekt og objekt går objektet fra å være nøytralt til å bli meningsbærende, der ulike egenskaper kan tilegnes et objekt. Disse egenskapene kan endre seg over tid, i ulike situasjoner og fra individ til individ, noe som vil si at det ikke er de fysiske egenskapene i seg selv som sier noe om hva objektet er. Det er *perceptual cues* som innprintes på objektet, av et subjekt det er i relasjon med, som blir definerende. For å ta et eksempel, kan en pennes kvaliteter gå fra å være et skriveredskap til et drapsvåpen ut i fra subjektets mentale tilstand, der objektet kan innta ulike *functional tones* i subjektets virkelighet.

Oppgave 9. What we talk about when we talk about context

Dourish viser til den økende innflytelsen såkalt *ubiquitous computing* har i mange aspekter av livene våres, der teknologien i større grad er innebygget i omgivelsene vi beveger oss i. Dette gjør at det tradisjonelle fokuset i HCI på interaksjon mellom teknologi og mennesket begrenset til et bestemt sted og situasjon ikke lenger er like relevant. Mennesker interagerer ikke lenger bare med teknologi som en enkelt bruker foran en stasjonær PC, men er noe vi interagerer med og har med oss på tvers av mange ulike kontekster. Innenfor HCI har det derfor blitt et større fokus på det Dourish (2003) referer til som ”context-aware computing” – som i større grad har en oppfattelse av konteksten den befinner seg i og kan respondere i henhold til denne. Dette gir opphav til nye spørsmål rundt hva kontekst er og hvilken rolle det spiller i interaktive systemer, noe Dourish undersøker i sin artikkel. Han argumenterer at tradisjonelle definisjoner av kontekst i litteraturen feiltoker rollen kontekst har i menneskelig aktivitet og foreslår en alternativ modell.

Dourish (2003) redegjør for relasjonen mellom sosiale og teknologiske aspekter, og da spesielt mismatchen mellom dem, som er et klassisk tema som behandles både i samfunnsvitenskapene og i computer science. Det som skiller disse to, er typisk det samme som skiller det positivistisk og fenomenologisk perspektiv på den sosiale virkeligheten generelt og relasjonen mellom teknologi og menneske spesielt. – som gjør at de har en veldig ulik oppfattelse av hva kontekst er.

Fra et positivistisk ståsted: kontekst er noe stabilt og forhåndsdefinert som kan reduseres til et sett med informasjon som kan representeres/modelleres i teknologi. I tillegg anser de aktivitet og kontekst som to separate ting. – mye forskning på ubiquitous computing baserer seg på en slik oppfatning.

Fra fenomenologisk perspektiv / Dourish sitt perspektiv: Kontekst er noe relasjonelt, noe som oppstår som situert og som aktivt produsertes i eksempelvis en bestemt setting eller aktivitet, og derfor heller ikke separat fra aktiviteten. Kontekst er derfor ikke et problem om representasjon men et interaksjonelt problem. Han foreslår derfor en interkasjonell modell av kontekst. Kontekst er en egenskap ved interaksjonen som oppstår i det som gjøres i øyeblikket. Eksempelvis oppstår og aktivt produseres konteksten av det å spille kort, hvem som spiller, hvor de spiller, hvordan de spiller osv. Dette kaller han for *embodied interaction*, der brukere forhandler og utvikler systemer av praksis og mening i interaksjon med systemer.

Oppgave 10. Black Mirror – ”Be Right Back” I denne episoden har en AI klart å kopiere minner, sjargong, stemme osv. til den avdøde ektemannen til hovedkarakteren. Hun opplever det etterhvert som om hun faktisk kan holde kontakten og ha et emosjonelt bånd til teknologien tilsvarende det hun hadde med ektemannen.

Oppgave 11. For et menneske handler kan det handle om fysisk og kognitiv frihet, og det å ikke være underlagt andres kontroll. Som setter premissene og reglene. Man er alltid autonom innenfor visse rammer, så man kan si at det finnes ulike grader for autonomi. Dette gjelder også for maskiner. Med større grad av autonomi jo større grad av premisser og regler defineres av agenten selv.

Oppgave 12. Begrepet Artificial intelligens, eller kunstig intelligens, ble first brukt av John

McCarthy I 1956 (Russel & Norvig, 1995).

Oppgave 13. Spørsmål til Dourish (2013): Hvorfor kan det være viktig at AI-systemer viser brukeren konteksten for valg den gjør, og hvordan påvirker det brukerens forståelse av egen adferd med systemet?

Oppgave 14. Spørsmål til Norman (1990): 'Hvilke situasjoner kan kreve det motsatte av en "soft, compliant technology," der man trenger at brukeren ikke undervurderer systemets evner og egenskaper?

15. Les "Like Having a Really Bad PA" by Luger & Sellen. Summarize in your own words key lessons learnt for interaction design with dialogue systems. Discuss the relevance of these lessons learnt for interaction with AI-based systems in general (1/2-1 page).

1. Å sette realistiske forventninger støtter opp om læringsprosessen

Luger og Sellen (2016) fant i sin studie, en tydelig mismatch mellom forventninger og den praktiske realiteten av bruk. For å forklare dette tok de utgangspunkt i Norman sitt konsept om 'gulf of execution and expectation'. Som tilsier at hvis representasjoner i systemet ikke enkelt kan oppfattes og dekodet til korrekte forventninger for bruk, vil det føre til en dårligere brukeropplevelse. Utfordringer ved å forstå CA sin evner gjaldt samtlige deltakere, men gapet mellom forventninger og realitet var enda større hos de med mindre teknologisk kunnskap. Deltakerne hadde i hovedsak to måter å tilnærme seg systemet på bakgrunn av sin manglende forståelse av dens evner: tilskrive den antropomorfe evner den ikke har, eller gjøre interaksjonen så enkel og 'økonomisk' som mulig. Sistnevnte gav en mer effektiv bruk og gjennomføring av oppgaver, men stod i veien for mer meningsfull bruk av CA. Førstnevnte førte ofte til frustrasjon

De med høy grad av teknologisk kunnskap har ofte andre mentale modeller og opplever interaksjonen ulikt. Eksempelvis en mindre tendens til å tilskrive CA *anthropomorphic* adferd, i tillegg til at de var mer tilgivende når CA feilet. De med lavere teknologisk kunnskap hadde større forventninger til systemet sinleggiens, som førte til større frustrasjon.

2. System-feedback og representasjon av intelligens:

Til tross for at noen brukere hadde en veldig økonomisk tilnærming til systemet og at det noen ganger feilet, var fortsatt forventningene til systemets intelligens høyere enn realiteten. Dette kan ha med folks forventninger til det å ha en samtale med noe. Eksempelvis at systemet brukte humor, som gjorde at noen opplevde at den utviste en viss sosial intelligens utover systemets faktiske evner. – leken interaksjon var viktig for å gjøre folk interessert, og var for nesten alle slik deres engasjement og eksperimentering med systemet startet, men gir det også en mer menneskelig karakter som påvirker forventninger.

3. Søtte opp om bedre evaluering av systemet – der systemets evner blir synlig:

Ofta gjør vi oss opp et bilde av en person og dens intelligens gjennom interaksjon, sosiale hint - hva vet, tenker og føler personen. I CA kan evner og begrensinger ofte

bli synlig for oss først når systemet feiler. De foreslår at man utforsker andre måter å gjøre dette synlig på.

4. Støtte opp om en pågående bruk ved å tydelig definere systemets formål:

Fokus på CA var sjeldent beskrevet som hovedaktiviteten når den var i bruk, noe som gjør CA *a means to an end*, heller en å være målet i seg selv. De argumenterer at design i dag ikke reflekterer det mest sentrale use case, og stiller spørsmål ved hvilke formål for bruk CA er designet for og at dette bør komme tydeligere frem i brukeropplevelsen. s

16. Beskriv med dine egne ord hva du forstår som ulike nivåer av automatisering.

- Hva er fordelene/ulempene relatert til høy/lav grad av automatisering (1/2 – 1 side)

I denne oppgaven har jeg tatt utgangspunkt i Sheridan og Verplank (1978), som skiller mellom 10 nivåer av automatisering, som spenner mellom tilefellet hvor mennesket har all kontroll til at maskinen har all kontroll. Nedenfor har jeg beskrevet noen av nivåene.

Det første nivået gjelder når det ikke forekommer noen grad av automatisering, og det er mennesket som tar alle beslutninger uten noen form for hjelp fra en maskin. Det andre nivået beskriver den laveste graden av automatisering ved at teknologien har en veiledende rolle i å vise eller skjule informasjon etter hva den anser som relevant for brukeren. Ved det femte, og midterste nivået, har mennesket fortsatt definert oppgavene men de utføres av maskinen. Det sjette nivået beskriver en beslutningsprosess der både mennesket og maskin kan foreslå mulige valgalternativer, men der det er mennesket som tar valget mellom dem og utfører dem med veiledning fra maskinen. På neste nivået, altså det syvende, er det fortsatt mennesket som velger mellom alternativer men her er det kun maskinen som definerer dem samt utfører dem. Det åttende nivået gjelder når maskinen definerer valgalternativer og velger mellom dem uten å forhøre seg med et menneske, foruten om mennesket ber den om det. Det nest siste nivået, nummer 9, beskriver en liknende situasjon men hvor maskinen informerer mennesket kun når den selv bestemmer seg for det. Hvis automatiseringen har nådd det siste nivået vil det si at maskinen agerer helt uavhengig av et menneske.

En høy grad av automatisering kan være fordelaktig når maskinen kan gjennomføre en oppgave mye mer effektivt og presist enn et menneske. Dette gjelder typisk for enklere oppgaver, som er repetitive i sin natur, slik som å fylle flasker med brus på en fabrikk. I mange tilfeller vil det å benytte en grad av automatisering (eksempelvis nivå seks) gjøre graden av menneskelig feil mindre. Et eksempel på dette kan være ved diagnostisering av kreft. Ved mer komplekse oppgaver som bedre utføres av et menneske, kan en for høy grad av automatisering være problematisk. Det samme gjelder også om maskinen må ta ta stilling til etiske problemstillinger, slik som hvem skal ofres i en evt. bil krasj, eller hvor/hvem en drone skal angripe.

Referanseliste:

- A. Følstad, (2018, 18. oktober). Interaction With AI- Module 2. Hentet fra <https://www.uio.no/studier/emner/matnat/ifi/IN5480/h18/undervisningsmateriale/interacting-with-ai---module-2---session-2---v03.pdf>
- Turing, A. M. (1950) Computing Machinery and Intelligence. *Mind* 49: 433-460.
- Domingos, P. (2012). A few useful things to know about machine learning. *Communications 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.
- Goertzel, B. & Pennachin, C. (2007). Artificial general intelligence (Vol. 2). New York: Springer.
- Google (2017) What is Machine Learning? Hentet fra <https://cloud.google.com/what-is-machine-learning/>
- Hutter, M. (2007) Universal Algorithmic Intelligence: A Mathematical Top->Down Approach. In Goertzel, B. & Pennachin, C. (Ed.), Artificial general intelligence (Vol. 2). New York: Springer. (p. 227-290)
- Lego Robotics (2015) Defining Robots and Robotics. Hentet fra <http://www.leorobotics.nl/definition-robots-and-robotics>
- Luger, E., & Sellen, A. (2016, May). Like having a really bad PA: the gulf between user expectation and experience of conversational agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 5286-5297). ACM.
- Norman, D. A. (1990). The 'problem' with automation: inappropriate feedback and interaction, not 'over-automation.' *Phil. Trans. R. Soc. Lond. B* 327(1241), 585-593.
- Russell, S. & Norvig, P. (1995). Artificial Intelligence: A modern approach. Prentice-Hall: Englewood Cliffs.
- Smith, C., McGuire, B., Huang, T., & Yang, G. (2006). The history of artificial intelligence. Seattle, Washington: University of Washington.
- Sheridan, T. B., & Verplank, W. L. (1978). *Human and computer control of undersea teleoperators*. MASSACHUSETTS INST OF TECH CAMBRIDGE MAN-MACHINE SYSTEMS LAB.
- Siciliano, B., Sciavicco, L., Villani, L., & Oriolo, G. (2010). Robotics: modelling, planning and control. Springer Science & Business Media.
- Susi, T., & Ziemke, T. (2005). On the subject of objects: Four views on object perception and tool use. *tripleC: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society*, 3(2), 6-19.
- Wikipedia (2015) Robotikk. Hentet fra <https://no.wikipedia.org/wiki/Robotikk>
- Wikipedia (2018) Maskinl ring. Hentet fra [https://no.wikipedia.org/wiki/Maskinl ring](https://no.wikipedia.org/wiki/Maskinl%C3%A6ring)