benjamtr
Høst 2022

IN5620 - individuell innlevering

Modul 1

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

Historien om AI

Begrepet artificial intelligence ble først introdusert i en workshop skrevet av den amerikanske matematikeren og logikeren John McCarthy i 1956 (Grudin, 2009, s.49). Det hele bemerker seg etter beregninger også kjent som datamaskinen spilte en stor rolle i andre verdenskrig (Grudin, 2009, s.49). Her spilte matematikeren Alan Turing en viktig rolle da han i 1949 i London Times skrev følgende, «I do not see why [the computer] should not enter the human intellect, and eventually compete on equal terms." (Grudin, 2009, s.49). Turing stadfester her viktigheten av datamaskinen og hvordan den har potensiale til å skape det samme som et menneske og om ikke bedre dersom man hyller det for den maskinen den er.

Definisjoner for AI

Definisjon 1

John McCarthy definerer AI som følgende:

> "… the science and engineering of making intelligent machines." (McCarthy, 1956)

Definisjonen fra 1956 sier noe om potensiale AI har da en intelligent maskin kan skapes gjennom AI. Innen kunstig intelligens mottok John McCarthy i 1971 en Turing Award (Wikipedia, 2018).

Definisjon 2

Neste definisjon kommer fra artikkelen Does AI make PD obsolete? Exploring challenge Artificial Intelligence to participatory Design skrevet av Tone Bratteig og Guri Verne:

> "… is 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 & Verne, 2018, s.1-2).

Bratteig og Verne er begge en del av instituttet for informatikk på UiO. Deres definisjon legger mer vekt på å skape noe som ligner menneskets evner i motsetning til McCarthy som ikke nødvendigvis sammenligner det med mennesket.

Definisjon 3

Den tredje definisjonen er hentet fra store norske leksikon hvor Axel Tidemann definerer AI slik:

> «… er informasjonsteknologi som justerer sin egen aktivitet og derfor tilsynelatende framstår som intelligent.» (Tidemann, 2020).

I likhet med de forrige definisjonene fokuserer man på utviklingen av noe intelligent, men her legger man også vekt på selve læringen til maskinen og hvordan den utvikler seg selv på egenhånd.

Egen Definisjon

«Kunstig intelligens er studien av teknologi som tar til seg informasjon og bruker det videre til å skape noe på lik linje med mennesket.»

Basert på de tre definisjonene har jeg kommet fram til en definisjon som vektlegger ny input som en viktig del av dens utvikling mot å kunne jobbe sammen med mennesket.

Menneske autonomi og maskin autonomi

Menneske autonomi deler likheter med maskin autonomi ved at begge tar valg som man tenker er blitt gjort av egen fri vilje. Forskjellen kommer kanskje mer frem da maskin

autonomien er mer eller mindre kontrollert og utviklet for spesifikke formål av menneske. Man kan derfor diskutere hvor autonom maskinen er da den i utgangspunktet bare kjenner til begrenset informasjonen gitt av mennesket.

Artikkel

Jeg har valgt artikkelen The 'Problem' with Automation: Inappropriate Feedback and Interaction, not 'Over-Automation' av Don Arthur Norman. I artikkelen presenterer Norman problemet med designet av kunstig intelligens i teknologi som samarbeider med mennesket. Her er ikke problemet at for mye teknologi er selvstyrt, men heller interaksjonen mellom det automatiserte og mennesket. Han mener problemet vil oppstå i det øyeblikket maskinen oppfattes som mer intellektuell enn mennesket selv (Norman, 1990). I et slikt tilfelle kan HCI bli mindre viktig som kan resultere i et stivere maskinere som ikke er i stand til å ta de livstruende avgjørelsene som det trengs at mennesket bryter inn i eller samarbeider om (Norman, 1990).

Selskap som jobber med kunstig intelligens

Det moderne selskapet Tesla er det første jeg tenkte på når det kom til noen som aktivt jobber med kunstig intelligens i dag. Hovedsiden presenterer automatisering på arbeidsplassen da mye gjøres av maskiner. De har også en egen side med tittelen kunstig intelligens og autopilot. Introduksjonen lyder som følger «Vi utvikler og distribuerer autonomi i stor skala for bruk i biler, roboter og mer. Vi tror at en tilnærming basert på avansert AI for visjon og planlegging, støttet av effektiv bruk av inference-hardware, er den eneste måten å oppnå en generell løsning som gir ren selvkjøring og mer til.» (Tesla, u.å.). Denne type vektleggingen av kunstig intelligens presenteres mer som en tjeneste (hvis ikke som et mål) enn noe annet. Tesla jobber ikke med en type plattform eller produkt, men heller et bredt spekter som bistår i studien av kunstig intelligens.

Film

I filmen Free Guy av Shawn Levy visualiseres en selvstyrt utvikling innen kunstig intelligens i et «open world»-spill. Da utviklerne av spillet skaper autonome karakter med rom til å lære og utvikle klarer de aldri å forutsi at deres kunstige intelligens bryter gjennom deres egne spilleregler. Gjennom menneskelig interaksjon med spillet fremstilles kraften av kunstig intelligens da den tar til seg så mye menneskelige egenskaper på en måte hvor et liv skapes i teknologien.

1.2 Robots and AI systems

Ordet robot

Ordet robot har sin opprinnelse fra et tsjekkisk teaterstykke på 1920-tallet med sin slaviske betydning "tjenslighet", "tvangsarbeid" eller "arbeid" (Science Friday, 2001).

Definisjoner for robot

Definisjon 1:

Den første definisjonen er hentet fra store norske leksikon og er definert av Roger A. Søraa fra NTNU:

> "Robot er en maskin som kan utføre handlinger automatisk, og som ved hjelp av elektroniske «sanser» (sensorer) og «oppskrifter» på hva de skal gjøre (algoritmer) kan bearbeide inntrykk fra omgivelsene, som igjen påvirker hvordan disse oppgavene utføres." (Søraa, 2022).

Dette er en definisjon som vektlegger "sansene" til en maskin og en gitt "oppskrift" som bestemmer hva som skal skje basert på omgivelsene sensorene møter på. Et eksempel vil være en "oppskrift" som sier at roboten skal dukke eller hoppe basert på om sensoren møter på en benk eller et lavt tak.

Definisjon 2:

Neste definisjon er gitt av The Free Dictionary og lyder som følger:

"menneskelignende maskin" (The Free Dictionary, u.å.).

Definisjonen er kort og betoner kanskje mer det utseendemessige mer enn noe annet. Da jeg tenker på en robot vil jeg først og fremst tenke på en maskin med menneskelige trekk som armer, ansikt og bein som utfører en oppgave på lik linje med menneske.

Egen definisjon:

"Robot er en maskin med gitte instrukser som automatiserer menneskelige oppgaver"

Min definisjon beskriver en maskin som har et spesifisert formål. Det trenger først og fremst menneskelig instrukser før det letter på arbeidskraften ved å automatisere oppgaver som i utgangspunktet ville trengt mennesker.

Forholdet mellom AI og roboter

Kunstig intelligens har på mange måter et tett forhold til roboter. Mange ville kanskje tenkt at alle roboter har kunstig intelligens integrert i seg, men definisjonene viser at det ikke nødvendigvis er riktig. De deler begge fellestrekk ved at de automatiserer oppgaver, men skiller seg kanskje mest ut ved at roboter ikke nødvendigvis trenger å være autonome på samme måte som kunstig intelligens. Definisjonene om roboter viser at det også kan være en maskin som har fått instrukser til å gjøre en oppgave som egentlig er tiltenkt menneskelig arbeidskraft. Et eksempel vil være en robot som skrur en kork på flasker. Den har fått instrukser til hva den skal gjøre, men som ikke nødvendigvis "justerer sin egen aktivitet og derfor tilsynelatende fremstår som intelligent." (Tidemann, 2020).

Kuri

Kuri er en familievennlig robot designet for å tilpasse seg hjemme. Med et ansikt som ser ut til å ha to øyne beveger den seg rundt veldig likt som R2-D2 i Star Wars. Den lærer seg planløsningen og strukturen av huset og tilpasser seg det daglige livet i hjemmet. Den interagerer med menneske enten ved berøring eller stemmegjenkjenning. Fritt beveger den seg rundt huset med funksjoner som lar den kommunisere, spille musikk/podcast og filme øyeblikk i hjemmet (heykuri, u.å.).

1.3 Universal Design and AI systems

Universell design

"Universal Design is the design and composition of an environment so that it can be accessed, understood and used to the greatest extent possible by all people regardless of their age, size, ability or disability." (universaldesign.ie, u.å.)

Definisjonen beskriver universell design som et inkluderende design. Med dette tatt i betraktning designer man ikke bare for en brukergruppe, men alle. Designet skal være forstått og kunne brukes av fellesskapet med ulike kognitive evner.

Potensiale av kunstig intelligens

Kunstig intelligens har et stort potensiale når det kommer til å designe for flere kognitive evner og oppfatninger. Et eksempel vil være selvkjørende biler som har muligheten til å bidra til at flere brukergrupper kan ta i bruk bilen. Kunstig intelligens som forklarer et bilde gjennom tekst er et annet eksempel som bidrar til at mennesker med blant annet nedsatt syn på lik linje kan oppleve et bildet.

Som nevnt tidligere gjennom eksempler har kunstig intelligens et stort potensiale når det kommer til å inkludere mennesker med ulike kognitive evner. Selv når kunstig intelligens jobber med å inkludere alle som ikke tidligere kunne benyttet seg av teknologien kan jeg fortsatt se for meg situasjoner av ekskludering. Et eksempel vil være stemmegjenkjenning som er tiltenkt for alle brukergrupper. Mange vil kunne artikulere seg riktig slik at systemet oppfatter det som blir sagt, men i tilfeller med nedsatt taleevne kan det oppstå problemer og misforståelser.

In the WCAG 2.1 principles and in many of the Human AI-Interaction guidelines the concept "understand" and "understanding" is used. Explain briefly in what way you make sense of the concept "understand" and "understanding". Then address the question: Do machines understand?

Bruken av "understand" og "understanding" i WCAG 2.1 prinsippene og retningslinjene for menneskelig AI-interaksjon

Jeg forstår konseptene "understand" og "understanding" som at de er nødvendige da det motsatte kan resultere i blant annet misforståelser og frustrasjon. Dette går begge veier da det er viktig at systemet er lett forståelig for brukeren som skal benytte det. Systemet/maskinen må også kunne forstå input fra brukeren slik at denne dataen kan brukes på riktig måte. Om maskinene virkelig forstår føler jeg avhenger av hvilken type maskin det er. Altså om det er kunstig intelligent eller en maskin som bare spiller musikk. De har begge en form for forståelse, men den ene har ofte bare en skreddersydd algoritme som den følger avhengig av input. Er ikke sikker på om det kan kalles forståelse dersom den aldri tar i betraktning hva informasjonen virkelig er.

1.4 Guideline for Human-AI interaction

Retningslinje: 11 Make clear why the system did what it did.

Retningslinje nummer 11 forteller noe om forståelse og tilbakemelding da brukeren skal oppdateres med informasjon om hvorfor systemet tok en avgjørelse. Et eksempel vil være at systemet informerer brukeren med en beskjed om at den lagde en liste med forslag til middagsretter da den oppfattet brukerens søk på mat. Jeg har valgt å sammenligne retningslinjene for menneskelige AI-interaksjon med "Shneiderman's Eight Golden Rules" i HCI. Først og fremst ser man likheter ved at de begge legger vekt på viktigheten av forståelse. Videre kan man diskutere hvorvidt forståelsen kan defineres på samme måte. I HCI vil forståelsen mellom maskin og brukeren være mer i fokus enn i AI-interaksjonen der man heller definerer viktigheten av at maskinen klarer å forstå menneske.

Modul 2

2.1 Kjennetegn ved AI-infused systemer

Nøkkelegenskaper

Guidelines *for Human-AI interaction beskriver AI-infused systemer som "systems that have features harnessing AI capabilities that are directly exposed to the end user"* (Amershi et al., 2019, s. 1). På den måten bruker systemet kunstig intelligens til å produsere eller kommunisere noe direkte til brukeren av systemet. Amershi et al. kjennetegner dette med uforutsigbare resultater som kan oppleves forvirrende, støtende og ikke minst farlig (Amershi et al., 2019, s. 1). Dermed er uforutsigbare resultater og error som false-positive og false-negative vanlige karakteristikk ved AI-infused systemer (Amershi et al., 2019 s. 1-2).

Stadige fremskritt blir gjort blant AI-teknologi som gjør det utfordrende for designere å designe effektive AI-infused systeme. Dermed kan mangel på forståelse lede til forvirring og error hos brukerne. Artikkelen *Will You Accept an Imperfect AI?: Exploring Designs for Adjusting End-user Expectations of AI Systems* diskuterer karakteristikkene videre ved å nevne at AI-funksjonalitet ikke er perfekt og vil kunne oppføre seg uforutsigbart (Kocielnik et al., 2019, s. 2). I likhet med de to artiklene diskuterer også Yang et al. utfordringer med AI blant designere. AI er et landskap som kan være vanskelig å forstå seg på som deriblant gjør det til systemer som er vanskelig å designe, implementere og samarbeide med andre om (Yang et al., 2020, s.2-3).

Dream by wombo

Dream by Wombo er et AI-infused system som genererer AI kunst basert på den informasjonen brukeren gir systemet. Det første jeg legger merke til etter å ha prøvd systemet er at jeg ofte får utforutsigbare resultater i forhold til informasjonen jeg har sendt inn. Da jeg sendte inn informasjonen "Elon Musk dancing at the club" fikk jeg tilbake et portrettbilde av en mann som det var vanskelig å kjenne igjen. Dette kan oppleves som frustrerende da systemet ikke forstår input fra bruker og tolker det på egen måte som er mer til forvirring. Det var også tilfeller hvor jeg uoppfordret fikk tilbake nakenhet på AI-kunsten

før systemet oppdaget det og slettet det fra systemet. Selv om systemet oppdaget feilen hadde bildet rukket å være ute lenge nok til at det kan oppfattes som støtende for brukerne.

2.2 Menneskelig AI-interaksjonsdesign

Hovedpoeng fra Amershi et al. (2019) and Kocielnik et al. (2019)

I *Guidelines for Human-Interaction* presenterer de 18 design prinsipper for menneskelig AI interaksjon. Motivasjonen for det nye forslaget kommer fra en lang fortid med designere og utviklere som ikke helt har klart å lage gode løsninger for AI-infused systemer (Amershi et al., 2019, s. 2). Vanlig problemer som error og uforutsigbarheterer viser at mange av de tidligere prinsippene ikke nødvendigvis har en forståelse for menneskelig interaksjon med AI. Dermed tar artikkelen for seg to spørsmål når det kommer til designprinsippene: Er prinsippene relevante og er de tydelig nok for brukerne av systemet? (Amershi et al., 2019, s. 7).

Artikkelen til Kocielnik et al. (2019) tar for seg brukernes forventninger når det kommer til kunstig intelligente systemer. Først og fremst foreslår de tre teknikker som setter forventninger til AI-systemet. Den første skal bidra til å estimere nøyaktigheten på systemets opptreden, den andre skal hjelpe brukerne med å forstå hva systemet gjør og den siste lar brukeren justere systemets opptreden/nøyaktighet (Kocielnik et al., 2019, s.6-7). Et siste hovedpoeng fra artikkelen er at to systemer med lik nøyaktighet kan føre til ulik oppfatning av nøyaktighet og bruker akseptanse avhengig om det er innstilt til å unngå falsk-positiv eller falsk negativ error (Kocielnik et al., 2019, s.12).

To design retningslinjer sammenlignet med dream by wombo

To av designprinsippene som presenteres er “make clear why the system did what it did” og “learn from user from user behavior” (Amershi et al., 2019, s. 3). I forhold til det AI-infused systemet dream by wombo så avviker den fra det første prinsippet ved at den ikke til tilbyr forklaringer for valg den tar. Systemet lar brukeren komme med innspill for kunsten, men gir ingen form for tilbakemelding for hvorfor den lagde noe helt annet en input fra bruker. Mangel på tilbakemelding og kvalitet kan ha noe med at jeg bruker gratisversjonen uten en registrert bruker. Retningslinjen kan derimot være en motivasjon til å implementere

gjennomgående feedback til brukeren. Dette kan bidra til å minimere forvirrelser som eventuelt skulle oppstå og gir systemet muligheten til å forklare valg som er blitt tatt. Basert på at det er en gratisversjon er jeg heller ikke overrasket over at den ikke tilpasser seg etter tidligere interaksjoner med brukeren. Uten en registrert bruker vil den antageligvis ikke lagre informasjon om pågående interaksjon. Det er dermed forståelig at man ikke inkluderer et slikt designprinsipp i en gratisversjon da det i utgangspunktet ikke lagrer informasjon om brukeren som videre kan brukes til å forbedre brukeropplevelsen.

Bender et al. (2021)

I artikkelen *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?* diskuteres en økende størrelse på språkmodeller og potensielle farer som følger dens utvikling (Bender et al., 2021). Basert på problemstillingen viser det seg at veksten på språkmodeller koster miljøet mye i form av ikke-fornybare energikilder (Bender et al., 2021, s. 612). Ikke er dette bare skadelig for miljøet, men også kostbart. Her foreslår de å rapporterte kostnader og treningsdata, og muliggjøre egne evalueringer blant forskerne for å vurdere farer/fordeler. Størrelsen av språk modellene er også et problem i form av omfattende treningsdata. Bekymringen er hvordan størrelsen kan bidra i å absorbere skadelig data som kan kode språkmodellen til å ta til seg blant annet rasistiske og sexistiske syn (Bender et al., 2021, s. 613). Derfor anbefaler forfatterne å investere ressurser i dokumentering slik at man kan forsøke å unngå skadelig datasett da man forsøker å etterligne menneskelig språk oppførsel.

2.3 Chatbots

Utfordringer med design av chatbots

Det kan sies at chatbots er og blir den nye måten vi kommuniserer og interagerer med nettsider og applikasjoner, men med kunstig intelligens i bilde kan det by på noe utfordringer (Følstad & Brandtzæg, 2017). Chatbots baserer seg rundt samtaler med brukeren som skaper nye måter å se HCI på hvor man heller må fokusere på å designe et system som forstår hva brukeren vil i motsetning til å få brukeren til å forstå systemet (Følstad & Brandtzæg, 2017, s. 40). Det er heller ikke nok å fokusere på en type bruker da

chatboten må kunne ta flere situasjoner og brukere i betraktning i form av et større nettverk (Følstad & Brandtzæg, 2017, s. 41). Yang et al. (2020) forklarer også sentrale utfordringer blant designere som en mangel på forståelse for hva AI egentlig er og hvordan man kan implementere og designe for AI sine uforutsigbare svar (Yang et al., 2020, s.2-3). Dette underbygger kjennetegn som error og forvirrelser i interaksjon som ofte kan komme fra mangel på forståelse hos designere (Amershi et al., 2019, s. 1).

Retningslinjene 1 og 2 i Amershi et al. (2019)

Retningslinje 1 “make clear what the system can do” og retningslinje 2 “make clear how well the system can do what it can do” gjelder det initielle ved interaksjonen mellom bruker og det AI-infused systemet (Amershi et al., 2019, s. 3). Ved å følge disse retningslinjene opp med den gjeldende chatboten kan man opprettholde realistiske antagelser hos brukerne (Luger & Sellen, 2016, s. 5294). Med forventninger som er realistiske i forhold til hva chatboten er i stand til kan det også forhåpentligvis minimere eventuelle forvirrelser i interaksjonen.

Litteraturliste:

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). Guidelines for Human-AI Interaction. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–13. https://doi.org/10.1145/3290605.3300233

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: Short Papers, Situated Actions, Workshops and Tutorial - Volume 2*, 1–5. https://doi.org/10.1145/3210604.3210646

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. https://doi.org/10.1145/3442188.3445922

Curi. (2018). *Explore Kuri.* Hentet 13.09.2022 fra https://www.heykuri.com/explore-kuri/#feature-speakers

Følstad, A., & Brandtzæg, P. B. (2017). Chatbots and the new world of HCI. *Interactions*, *24*(4), 38–42. https://doi.org/10.1145/3085558

Grudin, J. (2009). AI and HCI: Two Fields Divided by a Common Focus. *AI Magazine*, *30*(4), 48. https://doi.org/10.1609/aimag.v30i4.2271

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. https://doi.org/10.1145/2858036.2858288

Kocielnik, R., Amershi, S., & Bennett, P. N. (2019). Will You Accept an Imperfect AI?: Exploring Designs for Adjusting End-user Expectations of AI Systems. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–14. https://doi.org/10.1145/3290605.3300641

Norman, D. A. (1990). The 'problem ' with automation: Inappropriate feedback and interaction, not 'over-automation'. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *327*(1241), 585–593. https://doi.org/10.1098/rstb.1990.0101

Søraa, R. A. (2022) Robot. Store norske leksikon. Hentet 13.09.2022 fra https://snl.no/robot

Science Friday. (2011, 22. april). *The Origin of The Word 'Robot'.* Hentet 13.09.2022 fra https://www.sciencefriday.com/segments/the-origin-of-the-word-robot/

Tidemann, A. (2020). Kunstig intelligens. *Store norske leksikon*. Hentet 12.09.2022 fra https://snl.no/kunstig_intelligens

Tesla. (u.å.). *Kunstig intelligens og autopilot.* Hentet 12.09.2022 fra https://www.tesla.com/no_NO/AI

The free Dictionary. (u.å.). *Robot.* Hentet 13.09.2022 fra https://no.thefreedictionary.com/robot

Universaldesign.ie. (u.å.). *What is Universal Design?* Hentet 13.09.2022 fra https://universaldesign.ie/what-is-universal-design/

Wikipedia. (2018, 12. august). *John McCarthy.* Hentet 12.09.2022 fra https://no.wikipedia.org/wiki/John_McCarthy

Yang, Q., Steinfeld, A., Rosé, C., & Zimmerman, J. (2020). Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–13. https://doi.org/10.1145/3313831.3376301