Individuell innlevering høst 2022

Elise Håøy

1 Modul

1.1 Konsepter, definisjoner og historien om interaksjon med AI og autonome systemer

AI er en forkortelse for Artifical Intellegence, som på norsk blir kalt Kunstig Intelligens. For å sette ord på hva AI er vil det lønne seg å starte med en definisjon. AI blir presentert som en personalisert tjeneste som vil svare på spørsmålene dine og gjøre livet ditt lettere (Tone Bratteteig, 2018). Den gjør det mulig å resonere om sine egne handlinger, kan læres opp og vil forstå dine behov.

Historisk sett var den første autonome roboten som kunne resonere om sine egne handlinger «Shakey the robot». Denne roboten ble utviklet av SRI Artificial Intelligence Center in Menlo Park, California. Dette forekom på slutten av 60- og starten av 70-tallet (Brooks, 1999). Likevel var det faktisk så tidlig som på 30-tallet en datapioner med navn Alan Turing som først lekte seg med feltet AI. Han skapte en abstrakt datamaskin og en skanner. Den beveget seg frem og tilbake basert på minnet, leste det den fant og derav skrev ytterligere symboler. Handlingene til skanneren ble diktert av et program med tilhørende instruksjoner som var lagret i minnet i form av symboler. Dette gjør at maskinen kunne operere, modifisere og forbedre sitt eget program. Dette ble kalt en Turing-maskin, og dette er forgjengeren til alle dagens moderne datamaskiner (Copeland, 2022).

1. «Kunstig intelligens (AI) er evnen til en datamaskin eller en robot styrt av en datamaskin til å utføre oppgaver som vanligvis gjøres av mennesker fordi de krever menneskelig intelligens og dømmekraft. Selv om det ikke er noen AI-er som kan utføre det store utvalget av oppgaver et vanlig menneske kan gjøre, kan noen AI-er matche mennesker i spesifikke oppgaver.» (Copeland, 2022).

2. «AI er et underfelt av informatikk rettet mot å spesifisere og lage datasystemer som etterligner menneskelig intelligens eller uttrykker rasjonell atferd, i den forstand at oppgaven vil kreve intelligens hvis den utføres av et menneske.» (Tone Bratteteig, 2018).

3. «AI i form av digitale assistenter eller "chatbots" og maskinlæring (ML), som er en gren av AI og informatikk som fokuserer på bruken av data og algoritmer for å imitere måten mennesker lærer på, og gradvis forbedre nøyaktigheten. Det kunne forstå lære, og deretter handle basert på denne informasjonen.» (Oosthuizen, 2022)

Basert på disse definisjonene ser jeg flere felles faktorer. Noe å legge merke til er at i de to øverste definisjonene nevnes det at AI-en er evnen eller et datasystem som lærer, imiterer eller utfører oppgaver som vanligvis krever menneskelig intelligens eller dømmekraft. Mens i den nederste definisjonen defineres det som bruken av dataen og algoritmene imiterer måten mennesker lærer på. Men alle tre definisjonene har til felles at det handler om at AI-en kan utføre oppgaver i likhet med mennesker, og muligens gjøre det enda mer nøyaktig.

Ved å lage min egen spesifiserte definisjon ønsker jeg å trekke frem at AI er en gren innenfor informatikk som legger vekt på å handle ved å bruke data og algoritmer til å kunne utføre oppgaver som er basert på intelligens, dømmekraft og rasjonell atferd. Evnene vil også kunne læres opp og gradvis bli mer og mer nøyaktig.

Dette baserer jeg på det jeg har lært og lest meg opp på og som gir mest mening for meg. Når jeg tenker på AI tenker jeg på en algoritme som sitter på så mye data at den kan utføre oppgaver basert på typiske menneskers intelligens, dømmekraft og rasjonelle atferd. Samtidig som den læres opp og blir mer nøyaktig basert på at den innhenter mer og mer data.

Menneskelig autonomi går ut på at man er selvstendig, selvstyrende og kan bestemme over egne handlinger. Eksempler på maskin autonomi er data-innsamling og behandling, beregningsoppgaver og grensesnittdesign. Og dette kan ha store mulige effekter på menneskelig autonomi (Arto Laitinen, 26 October 2021). Maskin autonomi vil ikke ha muligheten til å være respektfull og dette stiller igjen spørsmål ved etikken bak disse autonome systemene, det er den største forskjellen mellom autonome systemer og menneskelig autonomi. Ved at vi mennesker naturlig handler (som oftest) ved hjelp av respekt

vil et autonomt system som for eksempel skal vise reklame føre til økt manipulasjon eller politisk undertrykkelse for å nevne noe (Arto Laitinen, 26 October 2021). Eksempler på likheter er at både mennesker og maskiner kan handle autonomt, altså begge er selvstyrende.

At noe er automatisk har direkte betydning med at en prosess blir utført uten menneskelig påvirkning. Dette kan gjerne være en statisk handling hvor et eksempel kan være at TV-en slår seg av etter en gitt tid. Et annet eksempel på automatisering er cruise control mens et autonomt system ville vært at en bil hadde operert av seg selv uavhengig av mennesket på innsiden (Jones, u.d.). Likheter med dette er at begge opererer uavhengig av et menneske, men automatisk er mindre teknisk krevende, det har ikke noe med miljøet rundt å gjøre, mens det har autonome systemer. Mens et autonomt system eller hvis noe er autonomt så vil det tilsi at under et spesifikt miljø vil en handling automatisk kunne utføres selv med betydelig usikkerhet i miljøet (Jones, u.d.).

Artikkelen jeg har valgt å diskutere er «The Ethics of AI Ethics: An Evaluation of Guidelines» (Hagendorff, 2020). Denne setter søkelys ved hvordan AI utfordrere etiske prinsipper og med dette har det blitt laget en hel del med nye regler som teknologiutviklerne bør følge. Denne rapporten diskuterer dermed om disse etiske retningslinjene har noe faktisk innvirkning på menneskelig beslutningstagning. Her mener forfatter Thilo Hagendorff at det korte svaret er «nei».

Det som var interessant med denne artikkelen er at mange systemutviklere ser på etikken rundt AI som en slags «byrde» som en ekstra ting å ta hensyn til i tillegg til design, brukeropplevelse ol. Mens jeg mener det burde vært en naturlig del av utviklingsprosessen. Ikke bare det å se på etikken i etterkant for å se om det det skrider med etiske prinsipper og verdier.

Dette innebærer at formålene som AI-systemer utvikles og brukes til, ikke er i samsvar med samfunnsverdier eller grunnleggende rettigheter som velgjørelse, ikke-ondskap, rettferdighet og forklarbarhet (Hagendorff, 2020). Det gjøres en stor teknisk innsats for å oppfylle etiske mål innen ansvarlighet og forklarbar AI, rettferdighets- og diskrimineringsbevisst datautvinning, samt personvern (Hagendorff, 2020). Det som er synd er at brudd på etiske retningslinjer ikke gir noen konsekvenser, annet enn på brukerne av systemene. Så jeg er enig

med forfatter at retningslinjene ikke direkte hjelper, grunnet at konsekvensene for å ikke følge de ikke er til stede.

Huddly AS er et firma som bygger smarte konferansekameraer drevet av AI on the edge. På nettsiden sin beskriver de at måten de bruker AI på er å bruke bransjeledende smart framing for å levere sosialt inkluderende videomøter. Basert på det vil jeg si at de snakker om AI som en service som hjelper de med å levere et utstående og moderne produkt (Huddly, u.d.)

Disney filmen Wall-E handler om en robot med alle menneskelige følelser og handlingsevner. Samtidig er det masse andre roboter i denne filmen som menneskene bruker, for eksempel en «vaske robot» som støvsuger og vasker alle steder det blir møkkete. Det er portrettert slik at noen av robotene er litt redde for mennesker, de har bare en jobb og menneskene vet ikke nødvendigvis at robotene har følelser. Om vi noen gang kommer så langt at vi kan ha roboter som handler, tenker og føler er jeg usikker på.

1.2 Roboter og AI systmer

Ordet "robot" er et gammelt kirkeslavisk ord «robota» som betyr tjenestlighet, tvangsarbeid eller arbeid. Det var idéen til en tsjekkisk dramatiker, romanforfatter og journalist ved navn Karel Čapek (1880-1938) som introduserte det i sitt hitspill fra 1920, R.U.R., eller Rossums Universal Robots (Markel, u.d.).

1. «A machine that resembles a human and does mechanical, routine tasks on command."

2. «Any machine or mechanical device that operates automatically with humanlike skill.»

(Begge definisjonene er fra dictionary.com)

Disse definisjonene beskriver begge to at en robot ligner på et menneske eller er menneskelignende. Den første beskriver at den gjør mekaniske rutine oppgaver eller

kommandoer, noe som ligner mye på definisjon 2 som beskriver at en robot operere automatisk med menneskelige ferdigheter.

Selv er jeg enig i begge de definisjonene ovenfor, men jeg ville beskrevet en robot som: «en menneskelignende maskin som utfører menneskelige handlinger automatisk». Dette fordi det er en maskin som ofte ligner mer et menneske enn noe annet, ikke nødvendigvis i utseende, men ut ifra hvordan roboten utfører handlinger. Samt at jeg synes først det er en robot hvis den utfører handlinger automatisk eller basert på forespørsler. Ikke hvis den er statist/mekanisk uten spesielle funksjoner.

Jeg mener at det kan være forskjell på AI og en robot, men de kan også kombineres og bli et sammensatt «system». En AI er et som sagt tidligere en gren innenfor informatikk som legger vekt på å handle ved å bruke data og algoritmer til å kunne utføre oppgaver som er basert på intelligens, dømmekraft og rasjonell atferd. Evnene vil også kunne læres opp og gradvis bli mer og mer nøyaktig. Dette er jo noe mennesker også kan gjøre, og i og med at definisjonen til en robot er en menneskelignende maskin som utfører menneskelige handlinger automatisk. Så vil jeg si at Robot og AI kan gå forholdsvis hånd i hånd.

Her er et godt eksempel roboten fra «No Isolations» som er spesielt utviklet for langtidssyke barn som må være mye borte fra skolen og sosiale sammenkomster. AV1 en liten robot med kamera, mikrofon og høyttaler som styres via en app som man kan ha på enten mobil eller nettbrett. Den er utviklet slik at eleven kan se og delta fullstendig i undervisningen. AV1 har støtte for 4G, og kan tas med ut i friminutt, på klasseturer eller bli med i bursdagsselskap. (No isolation, u.d.) Dette gjør så man kan være med virtuelt fra hvor som helst og passer på at de barna som ikke har mulighet til å være med på alt likevel får den fullstendige skoleopplevelsen.

1.3 Universelt design og AI systemer

Please find and describe a definition of Universal Design. Explain this definition and how you understand what Universal Design is about with respect to inclusion.
The Centre for Universal Design definerer universell utforming og hensikten med konseptet på følgende måte: "Universell utforming og design er utforming av produkter og omgivelser

på en slik måte at de kan brukes av alle mennesker, i så stor utstrekning som mulig, uten behov for tilpassing og en spesiell utforming" (Aslaksen, Heggem, Bergh, & Bringa, 1997).

Hensikten med universelt design er at en nettside skal være tilgjengelige for alle, uavhengig av forutsetninger. Det handler om å inkludere alle, skape et design som er forståelig, inkluderende og ikke diskriminerende. Måter man kan gjøre dette på er å utnytte farger, former, symboler, typografi for å forsterke budskap. Legge til rette for funksjonsnedsettelser som dårlig syn og fargeblindhet. Samtidig som man må ha en klar navigasjon på nettsiden. Mennesker tolker informasjon på forskjellige måter, med forskjellige sanser og det er viktig å inkludere alle.

Det er mye potensial i forhold til AI angående respekt for menneskelig oppfatning, bevegelse og følelser. AI er jo ment for å kunne erstatte enkle menneskelige handlinger og «gjøre hverdagen enklere». Et eksempel på dette er en robotgressklipper, som blant annet kan hjelpe eldre ved at de kan bo lenger hjemme, noe som er en lykkefølelse. Andre eksempler for inkludering er teknologi som google home, som ofte blir brukt av eldre som kanskje er demente eller ikke ser så godt lenger, som kan spørre om hva slags dag det er, hva klokken er, hva slags vær ol. Et AI-system i forhold til bevegelighet er et konsept som ikke er helt ferdig utviklet enda, men da tenke jeg på selvkjørende biler. Når det konseptet forhåpentligvis kommer på markedet, vil det gjøre bevegeligheten til mange mye større.

Som nevnt ovenfor er det flere eksempler hvor AI kan inkludere mennesker, er for eksempel en chatbot, her er det muligheter for at en person enkelt kan beskrive et problem til en chatbot. Her vil chatboten mest sannsynlig hjelpe deg. Men hvis personen som får hjelp da har dysleksi eller synsvansker kan det ekskluderer en bruker ved at det ikke er funksjoner som «høytlesning» eller «forstørr skrift».

Å forså betyr å oppfatte en mening om noe. Mens forståelse har en definisjon som lyder som å begripe eller fatte noe. Som går ut på at en bruker oppfatter og forstår handlingene som skal bli gjort og forstår hvorfor man må gjøre handlingen.
Et eksempel på om maskin forstår, så kan det at en data er programmert til å gjøre en handling og gjør den helt riktig, men den har ikke noe direkte forståelse av hvorfor den gjøre det, fordi den ikke har menneskelig følelser.

1.4 Guideline for Human-AI interaction

Retningslinje 11: "Make clear why the system did what it did"
Hensikten med denne retningslinjen er at det er viktig for bruker å forstå hvorfor systemet gjorde som det gjorde. Et eksempel her er hvis du skifter passord må du ofte logge ut for at systemet skal funke igjen, eller at bruker får beskjed om at du nå blir logget ut så ikke bruker blir skremt eller får en dårlig brukeropplevelse.

Det er mange likheter med HCI retningslinjer og AI guidelines. Noen likheter er at man skal unngå feilmeldinger, streve etter at det skal være universelt utformet og at man skal bli informert om hva systemer gjør. Å lære av hva brukeren gjør er mer for AI retningslinjer enn HCI siden en AI lærer og kan gjøre bedre.

Modul 2

2.1 Karakteristiske trekk ved AI-tilførte systemer

For å definer hva AI-tilførte systemer gjør ønsker jeg å trekke frem en beskrivelse: "systemer som har funksjoner som utnytter AI-evner som er direkte eksponert for sluttbrukeren" (Amershi, et al., 2019)

Fire nøkkelegenskaper ved AI-tilførte systemer: læring, forbedring, black box og hvordan det er drevet av store datasett. For å forklare dette så kan vi sarte med å se på «læring». Ved at et AI-tilført system lærer brukeren hva den kan gjøre eller oppnå, og hvor bra systemet kan gjøre det (Følstad, 2022). Med forbedring menes det at det at et AI-tilført system vil gjøre uunngåelige feil, derfor er det viktig å designe for usikkerheten ved at dette kan oppstå. Da vil det være veldig viktig å legge til rette for at det skal være enkelt å redigere, avgrense eller gjenopprette noe når AI-en er feil (Følstad, 2022). Med black box menes at et AI-tilført system skal vise tydelig for bruker hvorfor systemet gjorde som det gjorde. Så her handler det om å designe så bruker forstår og får en forklaring på utførte handlinger av AI (Følstad, 2022). Siste nøkkelegenskap er som nevnt over; drevet av store datasett. Med dette menes det at for at en AI skal funke, bli bedre og lære av sine tidligere feil vil det være helt nødvendig at den lærer av brukeradferd. Rett og slett å personlig-gjøre brukeropplevelsen ved at AI-en lærer av handlingene deres over tid. (Følstad, 2022)

De egenskapene nevnt ovenfor kan man se i det AI-funderte systemet Youtube. Et eksempel her er at AI-en til Youtube lærer seg å «kjenne» hver enkelt bruker basert på videoene brukerne spiller av. Videre fra dette vil en bruker få forslag til liknende videoer fra samme skaper eller lignende tematikk. Med tanke på «forbedring» vil AI-en lære seg hva den gjorde feil ved at du for eksempel klikker på «dette likte jeg ikke» når du har fått forslag på en video som ikke er av interesse. Noen ganger vil en bruker få opp helt tilfeldige videoer altså uten relasjoner fra tidligere viste videoer, dette gir brukeren en black box effekt.

2.2 Chatbots / kommunikasjon brukergrensesnitt

En chatbot kan defineres som en maskinagent som tjener et naturlig språk-grensesnitt til data og tjenester gjennom tekst eller stemme (Følstad, 2022). En chatbot kan ha flere formål som blant annet kundeservice, assistent, guide eller den kan være en sosial relasjon (Følstad, 2022).
Design av chatbots og/eller AI-tilførte systemer er en utfordring. Designere har rapportert at det er vanskelig å bedrive en fullverdig teknolog-drevet design prosess siden det ofte er vanskelig å forstå hva AI-en kan og ikke kan, og ut ifra det vanskelig å identifisere aktivitetene den kan/vil bedrive (Yang, Steinfeld, Rose, & Zimmerman, 2020). Derfor vil det være en stor utfordring å designe en chatbot som er stadig er i utvikling og som lærer seg nye ting. Skal den kun vise tekstformat eller kan den vise både bilder og videoer i tillegg? Dette er noen av nøkkelutfordringer med chatbots, men med tiden vil det stadig bli mer og mer utfordringer knyttet til tematikken.

En nøkkelfaktor for suksess som Følstad og Brandztæg beskriver i rapporten sin «Chabots and the New World of HCI» er å forstå samtale prosesser. Dette er en utfordring knyttet til hvor godt chatbots kan støtte samtaleprosesser samtidig som det gir nyttig utgang. De mest avanserte botsene i dag er Google Assistant hvor den har en evne til å holde en samtaltråd gjennom flere trinn i en dialog. Men selv assistenten samtalen bryter til slutt ned og innspillet fra chatboten blir irrelevant (Følstad & Bae Brandtzæg, Chatbots and the New World of HCI, 2017).

En annen nøkkelutfordring som Følstad og Brandztæg beskriver er i forhold til det å ivareta etikk og personvern. De mener det er helt nødvendig for forsker og praktiskere å gi nødvendig veiledning om etikk og personvern i forhold til chatbots og brukergrensesnitt for naturlig språk. Med ny og avansert teknologi innebærer det også ny etikk og personvern implikasjoner (Følstad & Bae Brandtzæg, Chatbots and the New World of HCI, 2017).

2.3 G1 og G2

Rettningslinjene G1: “Make clear why the system did what it did” og G2: “Make clear how well the system can do what it can do” (Amershi, et al., 2019) kan begge være med på å “løse” utfordrignene ved design av chatbots. Det å gi bruker en full forståelse av hva systemet kan gjøre og hvordan vil være med på å øke forståeligheten hos bruker og dermed eventuelt senke forventningene. Med dette mener jeg at brukeren er orientert om aktivitetene til AI-en eller chatboten og dermed ikke vil forsøke å få den til å utnytte andre aktiviteter (Kocielnik, Amershi, & Bennett, 2019). Et eksempel på hvordan man kunne løst utfordringene med G2 er at man kan gi retningslinjer som bruker av chatboten burde følge, eks: bruk nøkkelord, skriv muntlig språk, enkle ord og korte setninger er noe som muligens vil minke sannsynligheten for feil eller misforståelser.

Bibliografi

Amershi, S., Vorvoreanu, M., Weld, D., Fourney, A., Nushi, B., Collisson, P., . . . Bennett, P. (2019). *Guidelines for Human-AI Interaction.* Glasgow: Microsoft.

Arto Laitinen, O. S. (26 October 2021). *AI Systems and Respect for Human Autonomy.* Tampere, Finland: Faculty of Social Sciences, Tampere University,.

Aslaksen, F., Heggem, E., Bergh, S., & Bringa, O. (1997). *Universell utforming Planlegging og design for alle.* Oslo: Rådet for funksjonshemmede .

Brooks, R. A. (1999). *The early history of the new AI.* Massachusetts Institute of Technology.

Copeland, B. (2022). *artificial intelligence.* The Editors of Encyclopaedia Britannica.

Følstad, A. (2022). Hentet fra https://www.uio.no/studier/emner/matnat/ifi/IN5620/h22/interacting-with-ai-2022---module-2---session-1---handout.pdf

Hagendorff, T. (2020). *Guidelines, The Ethics of AI Ethics: An Evaluation of.* Minds And Machines, Springer Link.

Huddly. (u.d.). Hentet fra Huddly: https://www.huddly.com/

Jones, S. (u.d.). *Automation, Autonomy and the Messy In-Between*. Hentet fra ISG : https://isg-one.com/articles/automation-autonomy-and-the-messy-in-between

Kocielnik, R., Amershi, S., & Bennett, P. N. (2019). *Will You Accept an Imperfect AI? Exploring Designs for Adjusting End-user Expectations of AI Systems.* Glasglow: Microsoft.

Markel, H. (u.d.). *Science Friday*. Hentet fra 'Robot', The origin of the word: https://www.sciencefriday.com/segments/the-origin-of-the-word-robot/

No isolation. (u.d.). *No isolation*. Hentet fra AV1: https://www.noisolation.com/no/av1

Oosthuizen, R. M. (2022). *THE FOURTH INDUSTRIAL REVOLUTION.* University of South Africa, Pretoria, South Africa.

Tone Bratteteig, G. V. (2018). *Does AI make PD obsolete? Exploring challenges from Artificial Intelligence to Participatory Design.* Universitetet i Oslo.

Yang, Q., Steinfeld, A., Rose, C., & Zimmerman, J. (2020). *Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design.* Pittsburgh PA, USA : Carnegie Mellon University.

Vedlegg 1

Etter jeg hadde lest tilbakemeldingene valgte jeg å gjøre som han foreslo i 1. wish og det var å legge til en definisjon utarbeidet av eksperter på Universelt Design.

Tilbakemelding til Elise vedrørende første oppgave i IN5620

Hei! Her er min tilbakemelding på oppgaven din. Jeg har valgt å lage en liten vri på den foreslåtte strukturen på tilbakemeldingen ved å legge til en ekstra "star" og "wish".

1st. star:

Jeg syns du beskrev hvordan AI kan defineres godt. Definisjonene du viste til var gode, og det var lett å forstå hvordan du kom frem til din egen definisjon basert på disse.
I tillegg belyste du hva som bør inkluderes for at noe skal kunne bli definert som et AI-system på en god måte.

2nd. star:

Jeg syns det var interessant å lese drøftingen din knyttet til Hagendorffs artikkel. Jeg er enig med deg i at etikken rundt AI burde fokuseres på i utviklingsprosessen, og at man bør legge fra seg tanken om at det er en ekstra byrde å ta hensyn til, rett og slett fordi det er så viktig. Det er også interessant å lese diskusjonen du trekker frem vedrørende hvem som bærer på ansvaret dersom et system ikke har blitt utviklet i henhold til etiske retningslinjer. Her skriver at det ikke følger noen konsekvenser for andre enn brukere. Jeg tenker det er riktig i liten skala, men dersom det er snakk om masseproduksjon av et produkt som skal nå til massene ville jeg tippet at situasjonen hadde blitt en annen. Her er et eksempel:
https://www.forbes.com/sites/roberthart/2022/05/23/clearview-ai-fined-94-million-in-uk-for-illegal-facial-recognition-database/

3rd. star:

Måten du har forklart hva Universell Design (UD) er virker som er en god fremstilling. Her ser det ut til at du har tatt med de viktigste aspektene som begrepet er ment å dekke.

1. wish:

Selv om jeg liker fremstillingen din av hva UD er, savner jeg likevel en god definisjon fra en tredjepart som har ekspertise i dette fagfeltet. Det hadde gjort at jeg ville fått mulighet til å sammenligne både ditt og deres perspektiv, og således se om det var noen likheter og forskjeller i hvordan dere ville forklart begrepet.