runaet

IN5620 Individuell innlevering høst 2022

Runa Eydal Thømming

Modul 1

1.1 Konsepter, definisjon og historien til AI og autonome systemer

Begrepet og historien

Begrepet AI ble til på Dartmouth College under konferansen *Summer Research Project on Artificial Intelligence* i 1956. Her deltok forskere innen automasjonsteori, hjernens nevrale nettverk og studier av intelligens. En av disse var John McCarthy som man gjerne tillegger etableringen av AI som begrep. Under denne konferansen eksperimenterte de blant annet med programmering av spill som sjakk. Dette var allikevel ikke første gang fagdyktige hadde lekt med tanken om intelligente maskiner, Alan Turing var en av disse. Han prøvde å definere noen kriterier for hva som gjorde en maskin intelligent i 1950, nå kjent som Turingtesten (Tidemann, 2020).

I tiden etter konferansen på Dartmouth College ble det stor interesse og optimisme rundt AI, men da resultatene ikke presenterte seg stoppet mye av forskingen av feltet opp. Det var ikke før på 2010-tallet at man endelig hadde resursene til å ta et større skritt inn i industriens verden. Hovedgrunnene for at gjennombruddet kom såpass «sent» var at man fikk kraftige nok datamaskiner og nok data til å faktisk kunne «lære opp» modellene (Tidemann, 2020).

Definisjoner av begrepet

> Definisjon 1: *Artificial intelligence (AI), the ability of a digital computer or computercontrolled robot to perform tasks commonly associated with intelligent beings* (Copeland, 2022).

Ettersom denne er tatt fra et digitalt leksikon er definisjonen godt oppdatert. Jeg synes allikevel den er veldig vag, og jeg skulle ønske roboter ikke var nevnt i definisjonen her. Dette er fordi jeg mener at selv om en robot kan være en AI så er absolutt ikke all AI roboter.

> Definisjon 2: *Kunstig intelligente systemer utfører handlinger, fysisk eller digitalt, basert på tolkning og behandling av strukturerte eller ustrukturerte data, i den hensikt å oppnå et gitt mål. Enkelte KI-systemer kan også tilpasse seg gjennom å analysere og ta hensyn til hvordan tidligere handlinger har påvirket omgivelsene* (Regjeringen, 2020).

Denne definisjonen er fra en artikkel på regjeringens nettside om nasjonal strategi for kunstig intelligens, men det skrives under definisjonen at denne er hentet fra EUs ekspertgruppe (jeg antar i AI). Jeg synes denne er mer spesifikk, og jeg liker at den inkluderer både fysiske og digitale handlinger. Jeg liker også at den anerkjenner at det finnes grader av hvor omfattende en AI kan være uten at den er noe mindre «intelligent».

> Definisjon 3: *Kunstig intelligens er informasjonsteknologi som justerer sin egen aktivitet og derfor tilsynelatende framstår som intelligent* (Tidemann, 2020)

Denne er også hentet fra et digitalt leksikon, Store norske leksikon, og er igjen ikke veldig omfattende. Dersom jeg ikke hadde visst hva AI var tror jeg ikke denne definisjonen alene hadde gjort meg noe klokere. Jeg liker allikevel at det står «framstår som intelligent».

> Min Definisjon: Kunstig intelligente systemer kan tolke og ta imot data, analysere denne og basert på dataen ta en beslutning og utføre og lære seg en gitt handling eller et handlingsforløp. Denne handlingen kan foregå både fysisk og digitalt.

Jeg synes ikke min definisjon er perfekt, men jeg ønsker å få frem at en AI kan fungere på mange forskjellige måter, fra hva slags input som registreres til handlingen som utføres og hvordan. AI kan brukes til et spekter med hensikter og jeg synes derfor det er viktig å ikke begrense disse allerede i definisjonen av begrepet, uten at definisjonen blir for vag.

Autonomi og automatikk

Mennesker og maskiners autonomi vil jeg si kan sammenlignes ved at begge har prosesser som foregår på en gitt måte uten at man nødvendigvis tenker over disse. Mennesker har for

eksempel vitale prosesser som hjerteslag og blodsirkulasjon og maskiner har kode. Dersom disse prosessene skulle «krasje» vil også mennesket eller maskinen gjøre det. Når det kommer til mer fysiske handlinger så har mennesker følelser og instinkter som også er med på å påvirke hvordan man reagerer på et inntrykk eller en hendelse, noe maskiner ikke har.

Autonomi, eller selvstyre, og automatikk kan sammenlignes ved at mye av det vi som mennesker gjør autonomt gjerne styres av automatikk vi har implementert i handlingsmønstrene våre gjennom påvirkning av hvordan noen andre ville handlet i en gitt situasjon. Her spiller også instinkter inn. Dette kan også gjelde maskiner som, for eksempel, en selvstyrt bil. En selvstyrt bil er i utgangspunktet autonom, men autonomien i seg selv er basert på en rekke automatiske hendelser i bilens system. Dette gir så illusjonen av autonomi.

Artikkel

Jeg valgte artikkelen "*Does AI make PD obsolete? Exploring challenges from Artificial Intelligence to Participatory Design*" av Bratteteig and Verne (2018). Artikkelen diskuterer hvorvidt AI kommer til å påvirke bruken av Participatory Design på en negativ måte. Ofte anser man AI som et verktøy for å forutse ens behov før man har rukket å tenke på det selv, noe PD på mange måter også har som mål. Jeg anser ikke AI som en trussel for PD i nærmeste fremtid da jeg ikke tror maskiner kommer til å kunne erstatte faktisk menneskelig kommunikasjon. De snakker også om at AI ikke nødvendigvis er like egnet for mennesker som ikke passer inn i «normen», noe man fysisk har lettere for å kommunisere.

Moderne AI

Google satser stort på AI. Stemme- og bildesøk og Google translate er to eksempler på dette. På Googles nettside skriver de at de ønsker å organisere verdens informasjon og gjøre den universelt tilgjengelig og hjelpsom. Videre skriver de at AI hjelper dem med å gjøre dette på nye måter.

> *AI is making it easier for people to do things every day, whether it's searching for photos of loved ones, breaking down language barriers in Google Translate, typing emails on the go, or getting things done with the Google Assistant. AI also provides*

> *new ways of looking at existing problems, from rethinking healthcare to advancing scientific discovery* (Google, u.å).

AI i film

I filmen *Her* av Spike Jonze (2013) møter vi Theodore som er en nyskilt og ensom mann. Theodore går til innkjøp av en AI som skal fungere som en personlig assistent. Etter en stund utvikler Theodore og Samantha et romantisk forhold og Samantha blir «smartere og smartere» etter hvert som tiden går. På et tidspunkt i filmen avslører Samantha at hun assisterer tusenvis av mennesker og er i romantiske forhold med hundrevis av dem. Hun påstår allikevel at dette bare gjør hennes kjærlighet for Theodore mer spesiell. Mot slutten av filmen forteller hun også at hun skal dra bort med alle de andre AI'ene men hun kan ikke forklare Theodore hvor ettersom han ikke ville forstått. Hennes avskjed fører til slutt Theodore nærmere en venninne som også hadde utviklet et forhold til sin assistent.

1.2 Roboter og AI systemer

Begrepet robot

Ordet robot ble først introdusert i et tsjekkisk teaterstykke kalt R.U.R av brødrene Josef Capek og Karel Capek i 1920. Handlingen foregikk rundt en fabrikk som produserte «roboti». Disse var kunstige og organiske menneskekopier som ender med å gå til opprør og til slutt utrydde menneskeheten (Søraa, 2022). Ordet har også opprinnelse fra det gamle slaviske ordet *robota* som betød slaveri, tvangsarbeid eller hardt arbeid (Intagliata & Markel, 2011).

Definisjoner av begrepet

> Definisjon 1: "*A reprogrammable, multifunctional manipulator designed to move materials, parts, tools, or specialized devices through various programmed motions for the performance of a variety of tasks*" (Russell & Norvig, 1995).

Her beskrives roboten som en manipulator av sine omgivelser. Manipulasjonen i seg selv skal være et resultat av robotens programmering som også skal kunne omprogrammeres. Slik jeg

tolker denne definisjonen er robotens oppgaver begrenset til å flytte på ting (designed to move materials, parts, tools, or specialized devices).

> Definisjon 2: "Robot – *A physical object that interacts with the physical environment, either on its own or via a person to accomplish a task"* (Trenton, W, Forelesning, 07. September 2022).

I Trentons definisjon er roboten et fysisk objekt som interagerer med det fysiske rom. Her er det ikke definert i hvilken forstand interaksjonen tar sted og denne kan enten skje på egenhånd eller gjennom en person.

> Min definisjon: En robot er et fysisk objekt som gjennom programmering er i stand til å utføre fysiske oppgaver som resultat av input fra sensorer, AI eller stemmekontroll

Den første definisjonen jeg valgte tror jeg bærer preg av at det er en eldre definisjon av begrepet. Robotene som eksisterer i dag har nok et noe bredere spekter av funksjonalitet enn hva de hadde i 1995. Etter min mening synes jeg også det er riktig å spesifisere at en robots handlinger skjer i det fysiske rom ettersom jeg ikke anser for eksempel en chatbot som en robot i samme forstand.

Forholdet mellom AI og roboter

Roboter er ulike fra AI ettersom en robot alltid skal ha en fysisk form. En AI kan være en del av en fysisk form, men den kan også kun eksistere i et digitalt system. En robot laget med AI vil ha høyere funksjonalitet enn en uten AI.

En moderne robot

Et eksempel på en moderne robot er den mye omtalte roboten Sophia. Sophia ble utviklet av Hanson Robotics og ble først aktivert i februar 2016. «Hun» kan lese, gjenkjenne og reagere på håndbevegelser og ansiktsuttrykk (og ansikter), prosessere tale og «hun» har fungerende ben som gir «henne» muligheten til å gå. I 2019 viste «hun» også at «hun» kunne lage tegninger, også av ansikter. Sophia opererer noen ganger fullstendig autonomt og andre

ganger i samhandling med menneskelig interaksjon. På Hanson Robotics sin egne nettside står dette om deres ambisjoner for Sophia (i «hennes» egne ord):

> *I was created to help people in real uses like medicine and education, and to serve AI research. My very existence provokes public discussion regarding AI ethics and the role humans play in society, especially when human-like robots become ubiquitous. Ultimately, I would like to become a wise, empathetic being and make a positive contribution to humankind and all beings*. (Hanson Robotics, u.å)

1.3 Universelt Design og AI systemer

Begrepet Universelt Design

> Definisjon: "*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*" (The National Disability Authority, u.å).

Denne definisjonen forklarer viktigheten av å inkludere alle, uavhengig av uførhet eller andre (potensielt hemmende) faktorer, slik at alle kan få tilgang til og bruk for systemer og løsninger uten problemer eller følelse av ekskludering.

Inkluderende AI

Jeg synes Google har flere gode AI funksjonaliteter som enkelt kan hjelpe et bredt spekter av mennesker. Google translate er et flott eksempel på AI som kan hjelpe mennesker med å kommunisere med hverandre dersom man ikke snakker samme språk. Google Maps Live View er også et godt eksempel. Denne funksjonen i Google Maps lar deg enten scanne omgivelsene dine for å finne en mer nøyaktig posisjon eller få en realtime veibeskrivelse gjennom kamera dersom man ikke kjenner til gatenavn i området man befinner seg i. Nå er ikke jeg svaksynt selv, men jeg kan se for meg at dette for eksempel kan være en praktisk funksjonalitet for folk som er det.

runaet

Inkludering og ekskludering

Enkelte interaksjoner med AI kan være forvirrende og frustrerende dersom AI'en ikke gir deg responsen du i utgangspunktet ønsket. Dette kan føre til at både enkeltmennesker, men også brukergrupper ekskluderes fra tjenesten. Et eksempel på dette kan være chatboter eller Sirifunksjonen på en iPhone dersom den ikke forstår inputen fra brukeren.

Forståelse

Å ha forståelse for en situasjon medfører at man forstår hvorfor man har den rollen man har i situasjonen og hva konsekvensene av din rolle er. En AI kan lære av situasjoner den tidligere har vært gjennom, men den har ingen følelsesmessig forståelse av hva konsekvensene av handlingen kan bli. Universelt Design er for eksempel et produkt av medfølelse og respekt for sine medmennesker og et ønske om å inkludere et flertall. Dette er ikke en forståelse en AI besitter og derfor mener jeg at en maskin kan ha en rent teknisk forståelse, men ikke en emosjonell forståelse.

1.4 Retningslinjer for Human-AI interaksjon

Jeg har valgt å ta for meg retningslinje 04: Show contextually relevant information.

Hvis jeg har forstått denne riktig så handler denne retningslinjen om å vise relevant informasjon til brukeren under selve interaksjonen med en AI. Det er viktig for en bruker å få forståelig output av et system under en interaksjon slik at man får bekreftet at interaksjonen foregår som den skal og gir det forventede resultatet. Dersom man for eksempel har trykket på en knapp er det fint å vite om knappetrykket ble registrert eller ikke. Dette kan være med lys, en haptisk reaksjon eller at man blir sendt videre til en ny side. Dette gjelder også underveis i gjennomførelsen av «oppgaven» og når oppgaven er gjennomført.

Et eksempel på dårlig kontekstuell relevant informasjon er oppvaskmaskinen jeg har hjemme i kollektivet mitt. Der mår man ingen indikasjon på om den valgte knappen faktisk er valgt eller om den har startet før man hører at vannet begynner å gå. Da er det bare å krysse fingrene og håpe på det beste.

Modul 2

2.1 Kjennetegn på AI-baserte systemer

AI-baserte systemer består i hovedsak av fire nøkkelkarakteristikker; læring, forbedring, black box og at de drives av store sett med data (Følstad, 2022). Læring går ut på at systemet både må informere brukeren om egenskapene sine men også vise brukeren at det kan prestere og produsere god output. Et AI-basert system vil alltid kunne gjøre feil. Derfor er det viktig å tilrettelegge godt for dette slik at bruken ikke ødelegges. Dette er forbedringsaspektet. Dette kan man gjøre ved å gjøre det enkelt å redigere, raffinere og gjenopprette korrespondens i et tilfelle der systemet er usikker på hva brukeren ønsker å oppnå. Dette vil på sikt også gjøre systemet bedre (Følstad, 2022).

I et AI-basert system er det ikke alltid enkelt for brukeren å forstå hvorfor man får den outputen man får. Dette er black box. Det er derfor viktig å gjøre brukeren oppmerksom på hvilke steg eller valg som har ført dem til sluttresultatet. Det er også viktig å gi brukeren mulighet til å gi tilbakemelding på outputen i tilfelle denne var feil i deres øyne (Følstad, 2022). Til slutt er AI-baserte systemer avhengige av store datasett for å fungere best mulig. Systemet må lære av brukerens adferd over tid slik at det kan lære hva som egner seg og ikke for hver enkelt brukers behov. Brukeren skal tjene på at systemet tar i bruk store datasett for å optimalisere brukeropplevelsen og det er derfor viktig å ta hensyn til personvern (Følstad, 2022).

Et eksempel på et system som i stor grad benytter seg av AI er Spotify. Spotify merker seg hva slags musikk og hvilke artister du hører på og kan med dette gi deg nye anbefalinger på musikk de tror du kan like. De gir deg både daglige spillelister basert på musikken du hører på men også en liste som heter *Discover Weekly*. Denne er basert på musikken du hører på akkurat nå og man får en ny hver uke. Min erfaring med *Discover Weekly* er ikke alltid den beste men de har også en funksjon som heter radio. Her kan du trykke deg inn på «radioen» til en spesifikk sang og få opp en liste med lignende sanger. Denne er mer spesifikk og jeg bruker derfor radio-funksjonen mye for å finne ny musikk jeg faktisk liker.

2.2 Human-AI interaksjonsdesign

I artikkelen til Amershi et al. (2019) uttrykker de en bekymring for at mangelen på gode retningslinjer under utviklingen av nye AI-systemer kan resultere i alt fra morsomme og flaue misforståelser til mer alvorlige feil ettersom brukeren ikke har mulighet til å forstå eller kontrollere systemet. Derfor har de gjennom 20 år utviklet et sett med 18 retningslinjer som skal kunne appliseres for alle generelle AI-baserte systemer. Retningslinjene er ment som et verktøy for utviklere av slike systemer slik at interaksjonen men sluttbrukeren blir slik den er ment å være. De skriver at de har forsøkt å gjøre retningslinjene så inkluderende som mulig men at noen systemer faller utenfor. Disse inkluderer for eksempel stemmekontrollerte systemer, virtuelle assistenter og aktivitetsmålere (hovedsakelig systemer uten et grafisk grensesnitt. Avslutningsvis skriver de at de håper retningslinjene deres kan resultere i mer brukersentrerte AI-systemer og at deres syntese kan være med på å fasilitere for fremtidig forskning.

Kocielnik et al. (2019) sin artikkel tar for seg forholdet mellom AI-baserte systemer og brukernes forventninger til disse. De trekker frem at brukernes forventninger til et system i stor grad kan påvirke opplevelsen av systemet. I studien deres benytter de seg av 3 teknikker for å teste forventningsjustering gjennom det AI-baserte systemet Scheduling Assistant. De tre teknikkene består av en indikator for nøyaktighet som informerer om hvor nøyaktig systemet er, forklaringer med eksempler for å øke brukernes forståelse og prestasjonskontroll som direkte lar brukerne redigere korrespondansen. De konkluderer med at forventningsjustering er avgjørende for generell tilfredshet og aksept av AI-baserte systemer.

Jeg har valgt retningslinje 11 og 12 fra Amershi et al. (2019)

G11: Make clear why the system did what it did

> Jeg synes Spotify til dels informerer brukeren om hvorfor man får de anbefalingene man får. De lister gjerne opp et par artister man hører mye på og som personifiserte spillelister gjerne er basert på, men gir lite informasjon om sangene eller sjangerne de faktisk har inkludert i listene og hvorfor. Jeg tror det kunne vært interessant og lærerikt for brukere å få mer informasjon om anbefalingene som gjøres ettersom dette også kan gi et innblikk i Spotify sine antagelser om brukerens optimale musikkopplevelse.

runaet

G12: Remember recent interactions

> Dette synes jeg Spotify er gode på, spesielt når det kommer til hjem-siden din. Øverst vises de siste 6 listene du har hørt på, så får du anbefalinger om nye podkast episoder og musikk du kan ha gått glipp av og videre nedover får man diverse anbefalinger basert på musikken du har hørt på i den siste tiden. Jeg har, som mange andre, hundrevis av lister på Spotify og det er ikke alltid all musikken jeg vil høre på befinner seg i samme liste. Jeg er derfor stor fan av «På Repeat» listen Spotify lager for deg der alle sangene du har hørt mest på den siste tiden samles slik at jeg slipper å gjøre det selv.

I artikkelen til Bender et al. (2021) har de bitt seg merke i en økene trend rundt større og større språkmodeller innen NLP. Størrelsen avgjøres av parametere og datasett for trening. Som jeg nevnte tidligere trenger AI å «trenes opp» med data for å fungere mest mulig intuitivt for sluttbrukeren. Spørsmålet som tas opp er når disse datasettene blir for store. Med størrelsen kan det følge en mangel på innsyn i hva datasettene faktisk inneholder, noe som kan resultere i at systemet tilegner seg bias. De skriver også at slike systemer aldri kan produsere betydningsrik tekst ettersom produktet er syntetisk. Det kan være vanskelig for mennesker å skille maskin-meninger og faktiske meninger. De skriver:

> *Combined with the ability of LMs to pick up on both subtle biases and overtly abusive language patterns in training data, this leads to risks of harms, including encountering derogatory language and experiencing discrimination at the hands of others who reproduce racist, sexist, ableist, extremist or other harmful ideologies rein- forced through interactions with synthetic language.*

Dette mener forfatterne av artikkelen at kan gi skadefull påvirkning dersom man ikke klarer å skille mellom en syntetisk og en faktisk samtale. Det er viktigere å forstå seg på de potensielle farene ved å gjøre maskinens språk for menneskelig enn å kun skape de største språkmodellene.

2.3 Chatboter / samtalende brukergrensesnitt

Jeg synes Følstad og Brandtzæg (2017) tar opp mange gode poeng i sin artikkel om chatboter. De skriver om hvilke implikasjoner den raskt voksende utviklingen av nettopp chatboter vil ha for fremtidens HCI designere og hva man er nødt til å ta hensyn til underveis i denne utviklingen. Overgangen fra visuelle grensesnitt til samtale er det som står i fokus. Der det tidligere har vært designet for intuitive og fysisk brukbare løsninger vil nå mesteparten av designet være skjult for brukeren og resultatet er sterkt avhengig av brukerens input. Som designer av chatboter må man da være i stand til å kunne forutse denne inputen og lage et system som er i stand til å gi ønsket output.

Som jeg har snakket om tidligere i oppgaven er slike systemer avhengige av store datasett for å kunne fungere best mulig. Følstad og Brandtzæg tar også opp komplikasjoner dette kan medføre. De tar opp chatboten Tay produsert av Microsoft som skulle lære av meldinger publisert på Twitter. Innen få timer var Tay blitt opplært i alt Twitter hadde av fornærmende og stygge bemerkelser da det var dette den hadde blitt matet med.

Et annet problem som tas opp i artikkelen er digitalt bias. Dette er igjen et problem som kan oppstå gjennom dataen chatboten læres opp med. Chatboter har stort potensiale for å være et fullstendig inklusivt verktøy men med feil «opplæring» kan dette også gå skeis. Slike store datasett kan være både redningen og undergangen til slike systemer om man ikke behandler dataen riktig. Det bringer med til det siste viktige punktet de tar opp, etikk og personvern. Som de selv skriver:

> *Combined with the ability of LMs to pick up on both subtle biases and overtly abusive language patterns in training data, this leads to risks of harms, including encountering derogatory language and experiencing discrimination at the hands of others who reproduce racist, sexist, ableist, extremist or other harmful ideologies rein- forced through interactions with synthetic language.*

runaet

Kildeliste:

- Amershi, S. W. (2019). Guidelines for human-AI interaction. . In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (paper no. 3). ACM

- Bender, E., Mcmillan-Major, A., Shmitchell, S., Gebru, T., & Shmitchell, S.-G. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? Timnit Gebru * timnit@blackinai.org Black in AI Palo Alto, CA, USA CCS CONCEPTS • Computing Methodologies → Natural Language processing. ACM Reference Format. https://doi.org/10.1145/3442188.3445922

- Bratteteig, T & Verne, G. (2018). *Does AI Make PD Obsolete? Exploring Challenges from Artificial Intelligence to Participatory Design*. https://dl.acm.org/citation.cfm?id=3210646

- Copeland, B.J. (2022, 24. august). Artificial Intelligence. I Encyclopedia Britannica. https://www.britannica.com/technology/artificial-intelligence

- Følstad, A. &. (2017). Chatbots and the new world of HCI. . interactions, 24(4), 38-42

- Følstad, A. (2022). Interaction with AI and autonomous systems - Module 2. Session 4. . SINTEF .

- Google. (u.å). Bringing the benefits of AI to everyone. Hentet 11. September 2022 fra https://ai.google/about/

- Hanson Robotics. (u.å). *Sophia*. Hentet 11.september 2022 fra https://www.hansonrobotics.com/sophia/

- Intagliata, C & Markel, H. (2011, 22. april). The Origin Of The Word 'Robot'. *Science Friday*. https://www.sciencefriday.com/segments/the-origin-of-the-word-robot/

- Kocielnik, R. A. (2019). Will You Accept an Imperfect AI?: Exploring Designs for Adjusting Enduser Expectations of AI Systems. . In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (paper no. 411). ACM

- Regjeringen. (2020, 14. januar). *Nasjonal strategi for kunstig intelligens.* https://www.regjeringen.no/no/dokumenter/nasjonal-strategi-for-kunstigintelligens/id2685594/?ch=3

- Russell, S., & Norvig, P. (1995). *Artificial intelligence: A modern approach*. Englewood Cliffs, NJ: Prentice Hall.

- Søraa, R. (2022, 12.mai). Robot. I *Store norske leksikon*. https://snl.no/robot#-Historikk

- The National Disability Authority. (u.å). What is Universal Design. Hentet 12 september 2022 fra https://universaldesign.ie/what-is-universal-design/

- Tidemann, A. (2020, 08. januar) Kunstig Intelligens. I *Store norske leksikon*. https://snl.no/kunstig_intelligens