IN5620 Individuell oppgave høst 2022
Daniel Gloppestad Bajer

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

Det var John McCarthy som først brukte begrepet AI. Han var en del av en teknologioptimistisk bevegelse i etterkrigstiden, men hadde et annet syn enn mange fra sin samtid. Mange så for seg at AI om ikke lenge ville kunne handle intellektuelt på linje med mennesker. Med et slikt standpunkt så man naturligvis også for seg at AI etter hvert ville utføre oppgaver tidligere ansett som intellektuelle, mye bedre enn det mennesker var i stand til, og på den måten utkonkurrere mennesket. Alan Turing var en viktig person i den tidlige populariseringen av AI, og med sin velkjente Turing -test bidro han til ideen om at AI skulle kunne måle seg med mennesket. McCarthy mente på sin side at AI ikke nødvendigvis skulle sammenlignes med menneskelig intelligens og at AI var en del av Computer Science, ikke psykologi.

Science fiction litteraturen spilte også sin rolle med å gjøre kunstig intelligens til et kjent konsept og AI -relatert forskning og arbeid fikk mye økonomisk støtte på 40 og 50 -tallet. Man hadde sett datamaskinens potensial i praksis under andre verdenskrig, så det var store økonomiske krefter i sving for å støtte AI -utviklingen. *(Grudin, Jonathan. 2009)*

Det skulle etter hvert vise seg at AI ikke utviklet seg så fort som først spådd og selv i dag, hvor datamaskiner er i stand til å gjøre mange oppgaver bedre enn mennesker, er det vanskelig å argumentere for at AI skal kunne erstatte menneskelig intelligens.

Schuett, Jonas. (2019) presenterer tre definisjoner av AI.

1. AI er en datamaskin som består Turing -testen.
    - Alan Turing, 1950

2. AI betyr vitenskapen og oppbyggingen av intelligente maskiner. Intelligens refererer her til regningen som ligger bak evnen til å oppnå mål i den virkelige verden.
    - John McCarthy, 2007

3. AI er en intelligent agent, det vil si en maskin som har evnen til å ta inn sensorinput fra verden rundt seg og prosessere disse for å utføre handlinger som følger en gitt heuristikk.
    - Russel og Norvig, 2009

Av disse definisjonene er det McCarthy sin jeg ønsker å basere meg på i størst grad. I denne oppgaven vil jeg gi følgende definisjon av AI:

AI er en entitet som er stand til å samle inn og prosessere data for å oppnå forhåndsdefinerte mål.

Som mennesker er vi i stand til å ta egne valg. Eller vi lever hvert fall i illusjonen om at vi er det. En måte å se på menneskelig autonomi på er å tenke seg at vi er bundet til vår biologiske eksistens og at våre lyster og begjær ikke tillater oss å handle ut ifra egen autonomi. Mennesker kan likevel føle at de handler autonomisk og i denne oppgaven vil jeg gå ut fra denne definisjonen av menneskelig autonomi. Min definisjon av menneskelig autonomi er:

En følelse av å kunne gjøre egne valg og å handle på en måte som gir mening for enkeltmennesket og dets omgivelser.

Vi sammenligner denne definisjonen med hvordan autonomi defineres av *Antsaklis, & Rahnama, A. (2018)*

A system is autonomous with respect to a set of goals and
under a set of (subject to certain classes of) uncertainties,
if it achieves these goals under those uncertainties without
external intervention.

Her ser vi at følelser ikke er inkludert i *Antsaklis, & Rahnama, A. (2018)*, og det er vanskelig å tenke på maskiner som følende vesener. Man kan også se en forskjell mellom definisjonene der den første bruker ordet *mening*, bruker *Antsaklis, & Rahnama, A. (2018)* måloppnåelse. Man kan her argumentere for at mennesker trenger mening for å føle autonomi, mens maskiner ikke har noen følelse for mening, bare måloppnåelse. I tillegg definerer man et avgrenset system innenfor maskinautonomi, nettopp for å kunne kontrollere måloppnåelse og eventuell menneskelig kontroll utenfra. Menneskelig autonomi på den annen side er ikke adskilt fra påvirkning utenfra. Det er fullt mulig, og ganske naturlig, for et menneske å føle autonomi selv ved påvirkning fra omgivelsene. Dette er ikke en hindring for menneskelig autonomi fordi vi alltid er knyttet sammen med omgivelsene våre.

På *dictionary.com* finner vi følgende definisjon av hva *automatisk betyr:*

automatic
having the capability of starting, operating, moving, etc., independently:
an automatic sprinkler system; an automatic car wash.

Vi kan se at i likhet med definisjonen for *autonomi* finner vi også i denne definisjonen et fokus på uanvhengighet, altså noe som er stengt fra påvirkning utenfra. Det kan virke som om den største forskjellen mellom en autonom og en automatisk maskin, ut ifra definisjonene ovenfor, er at en autonom maskin er i stand til å håndtere en større grad av usikkerhet. Hvis vi sammenligner en vaskemaskin med en robotgressklipper, kan vi tenke oss at vaskemaskinen er automatisk mens robotgressklipperen er autonom. Vaskemaskinen utfører et bestemt program, mens robotgressklipperen er nødt til å gjøre vurderinger og valg underveis i jobben basert på skiftende omstendigheter. Hvis man ser nærmere på saken, kan man imidlertid se at det ikke er så enkelt å klassifisere inn i automatisk og autonom. En vaskemaskin må for eksempel tilpasse ytelsen sin basert på hvor mye klær som vaskes om

gangen. Dersom det er mye klær kreves det mer motorkraft for å oppnå et bestemt antall omdreininger i minuttet. Dette kan sees på som en form for autonomi. Man kan dermed argumentere for at forskjellen mellom en automatisk og autonom kun er basert på grad av kompleksitet, og det er vanskelig å sette en konkret grense for hva som er komplekst nok til å kunne regnes som autonomt, eller omvendt for automatisk.

Tone Bratteteig and Guri Verne. 2018 diskuterer om AI gjør Participatory Design (PD) overflødig. De peker på at moderne AI -teknologi deduserer sine egne regler basert på store mengder data. Dette skaper en utfordring for designere fordi det blir vanskelig å forstå valgene som gjøres av AI -teknologien som former de AI -skapte resultatene og produktene. Det er også vanskelig å forutse hvordan fremtidig teknologi skapt av AI vil fungere, ettersom det vil endre seg over tid, og dermed vanskelig å ta det med i betraktningen i en designprosess. Det er sentralt innen PD at konteksten og situasjonen skal være sentrale i designprosessen. Dette blir også en utfordring med tanke på at maskinlæring baserer seg på store mengder data som bidrar til generalisering og større avstand til den aktuelle konteksten. Hvis man imidlertid er bevisst hvilke data AI -systemet trenes på og sørger for at disse kommer fra en relevant kontekst har man gjort et viktig steg. De konkluderer med at AI står for en stor utfordring for PD, både for designere, deltakere, men at PD også kan være nyttig for å møte utfordringene som oppstår sammen med økende bruk av AI

Jeg er enig med synet som presenteres i denne artikkelen. Jeg mener det er viktig å sette fokus på brukeren i en tid der mer og mer automatiseres.

På sine hjemmesider presenterer *Meta* sin språksatsing på følgende måte:

No Language Left Behind (NLLB) is a first-of-its-kind, AI breakthrough project that open-sources models capable of delivering evaluated, high-quality translations directly between 200 languages—including low-resource languages like Asturian, Luganda, Urdu and more. It aims to give people the opportunity to access and share web content in their native language, and communicate with anyone, anywhere, regardless of their language preferences.

Meta presenterer her AI som kan oversette mellom 200 språk. De presenterer det som en tjeneste, som i tillegg er open source. Det kommer likevel også frem at de ser på denne teknologien som sentral i sin metaverse -ide, som også vil fungere som en verdensomspennende økologi.

Filmen Her av Spike Jonze fra 2013 handler om en mann som forelsker seg i et operativsystem. Filmens handling utspiller seg i en virkelighet der det er veldig vanlig å til enhver tid være tilkoblet en digital personlig assistent, og kommunisere med denne verbalt. AI presenteres som en allvitende entitet, en entitet som kan hente opp en enorm mengde informasjon. Denne interaksjonen med AI minner mye om hvordan vi forholder oss til smartassistenter som for eksempel Apples Siri i dag, men bruken er mye mer omfattende i filmen.

Utover i filmen utvikler hovedpersonens forhold med sin smartassistent seg til å bli romantisk og det ender opp med at smartassistenten dumper ham. Smartassistenten uttrykker at *hun*

trenger mer sofistikert stimulering og at hun ikke kan finne dette hos menneskelige partnere og derfor vil forsvinne inn i internett. I denne avsluttende delen av filmen møter vi på former for AI som foreløpig ikke eksisterer i virkeligheten. Den kunstige intelligensen viser menneskelige trekk som følelser, ambisjoner og behov for tilhørighet.

1.2 Robots and AI systems

I følge dictionary.com har ordet *robot* følgende opprinnelse.

Czech, coined by Karel Čapek in the play R.U.R. (1920) from the base robot-, as in robota compulsory labor, robotník peasant owing such labor

I Sebastian Thrun (2004) finner vi to definisjoner av hva en robot er.

In 1979, the Robot Institute of America
defined a robot as "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). In contrast, the Merriam Webster's collegiate dictionary (1993) defines
a robot as "An automatic device that performs functions normally ascribed to
humans or a machine in the form of a human."

Vi ser her at Russell og Norvig ser på konseptet robot som en maskin som kan utføre fysiske oppgaver. På den annen side ser vi at Webster definerer en robot som et stykke utstyr som kan utføre oppgaver som vanligvis utføres av mennesker. Hun sier ikke eksplisitt at en robot må gjøre fysiske oppgaver, men hun er nærmere en moderne definisjon av AI ved å peke på at en robot kan gjøre oppgaver som vanligvis utføres av mennesker. En smartassistent vil for eksempel være en robot etter Websters definisjon, men ikke etter Russell og Norvigs, siden den ikke utfører fysiske oppgaver.

I denne oppgaven vil jeg definere en robot som *en AI -maskin som er i stand til å utføre fysiske oppgaver*, i henhold til definisjonen av AI gitt i del 1.1.

Vi ser på en Tesla -bil som et eksempel, for å se hvordan denne passer inn i definisjonen av en robot, og for å se hvordan mennesker interagerer med den. Som en bil er en Tesla i stand til å utføre fysiske oppgaver. Hvis vi tar for oss en selvkjørende versjon av Teslaen, kan vi se at den også er et AI -system og dermed oppfyller begge kravene fra denne oppgavens definisjon av en robot. Grunnlaget for at vi kan kalle en Tesla for et AI -system er at den samler inn og prosesserer data, både gjennom egne sensorer og fra databaser for veitrafikkregler. I tillegg handler den i henhold til forhåndsdefinerte mål som for eksempel, å ikke krasje og å kjøre fra A til B. Den eneste inputen en Tesla krever av sin menneskelige bruker er A og B og så er den i stand til å gjøre resten selv. Den har imidlertid flere feedback og informasjonssystemer for at brukeren skal føle seg trygg. I tillegg finnes det også mulighet for at brukeren kan overstyre Teslaens handlinger.

1.3 Universal Design and AI systems

I Iwarsson, Susanne, and Agneta Ståhl finner vi følgende definisjon av Universell utforming, skrevet av Mace (1985)
Universal design is an approach to design that incorporates products as well as building features which, to the greatest extent possible, can be used by everyone

Denne definisjonen understreker at universelt design skal skape produkter som kan brukes av alle. Sentralt i fagfeltet Universell utforming finner vi tanken om at det skal designes for alle fra starten av design prosessen. Dette skiller seg fra eldre tankeganger om at systemer kan tilpasses brukere med spesielle behov i en avsluttende fase av systemutviklingen.

AI gir store muligheter når det kommer til å inkludere mennesker med spesielle behov. Hvis vi tar for oss produkter som kan hjelpe hørselhemmede og synshemmede, finner vi mange løsninger som baserer seg på AI -teknologi. AI kan blant annet hjelpe til med følgende oppgaver:

- Hjelp ved navigasjon og rutevalg
    - For eksempel Google Maps
- Hjelp med å identifisere fysiske objekter
    - For eksempel Seeing AI fra Microsoft
- Hjelp med å omforme verbal informasjon
    - Tale til tekst -teknologi

Disse teknologiene kan brukes for at flere mennesker skal føle seg inkludert i samfunnet, og også for å designe systemer med tanke på universell utforming.

Hvis vi ser på AI -teknologi for såkalt Augmented Reality (AR), er dette en bruk av AI som er med på å ekskludere mennesker som av en eller annen grunn ikke har tilgang eller mulighet til å bruke AR. Det kan være eldre mennesker som ikke har nok kunnskap eller ferdigheter til å bruke teknologien, eller for eksempel synshemmede som ikke har mulighet til å ta til seg visuell informasjon.

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?
You find the WCAG 2.1 principles here:
https://www.w3.org/TR/WCAG21/

Det er vanskelig å teste om en annen person eller maskin forstår et konsept. Derfor ønsker jeg i denne oppgaven å gå ut fra følgende definisjon av *forståelse*:

Forståelse er en følelse av at et konsept gir mening, eller at det passer med andre konsepter man har kjennskap til.

Ved denne definisjonen kan man argumentere for at maskiner kan forstå. Maskiner kan registrere data og sammenligne det med data som er registrert fra før og eventuelt se om det passer inn i et system eller trend. Dette minner om hvordan mennesker tar til seg ny

informasjon og forstår den.

Det er likevel forskjeller mellom hva mennesker og maskiner kan forstå. Maskiner kan for eksempel ikke forstå en tanke slik som mennesker, siden en tanke ikke kan representeres i sin helhet med diskrete data. Det er også en forskjell i hvordan mennesker og maskiner forstår, og vi kan finne noe av det i begrepene *følelser* og *mening.* Mennesker sammenligner på en annen måte enn maskiner for å finne ut om noe gir mening. Vi sammenligner ikke nødvendigvis konsepter mot andre konsepter for å finne mening. Følelsen av forståelse åpner også opp for grader av forståelse hos mennesker.

1.4 Guideline for Human-AI interaction

I Microsofts 18 retningslinjer for interaksjon med AI finner vi blant andre følgende retningslinje:

> *09 Support efficient correction*
> *Make it easy to edit, refine, or recover when the AI system is wrong.*

Her kan vi igjen se på Tesla som et eksempel. Tesla tillater brukeren å ta kontroll over bilen når brukeren selv ønsker det. Dette virker som en god løsning ettersom det ikke er sikkert at Teslaen selv registrerer at det har oppstått en feil.

Abascal, & Nicolle, C. (2005) presenterer HCI -retningslinjer for inkluderende design. En av retningslinjene sier at man skal minimere maskinautomatikk der det er mulig, slik at brukeren ikke mister følelsen av kontroll. Dette går veldig imot utviklingen i AI -teknologien i dagens samfunn og viser til noe av det samme problemet som *Tone Bratteteig and Guri Verne. 2018.* Utenom det har *Abascal, & Nicolle, C. (2005)* et fokus på etikk, personvern og menneskelig autonomi, mens Microsofts retningslinjer fokuserer mer på at brukeren skal holdes informert og forstå hvordan den kan interagere med systemer. Dette finner vi også igjen hos *Abascal, & Nicolle, C. (2005)* sine retningslinjer, men man kan argumentere for at de har et bredere syn, spesielt med tanke på etikk.

Kilder:

Grudin, Jonathan. "AI and HCI: two fields divided by a common focus." *AI Magazine*, vol. 30, no. 4, winter 2009, pp. 48+. *Gale Academic OneFile*, link.gale.com/apps/doc/A216267417/AONE?u=oslo&sid=bookmark-AONE&xid=d85837d1. Accessed 13 Sept. 2022.

Schuett, Jonas. (2019). A Legal Definition of AI. SSRN Electronic Journal. 10.2139/ssrn.3453632.

Antsaklis, & Rahnama, A. (2018). Control and Machine Intelligence for System Autonomy. Journal of Intelligent & Robotic Systems, 91(1), 23–34. https://doi.org/10.1007/s10846-018-0832-6

https://www.dictionary.com/browse/automatic

Tone Bratteteig and Guri Verne. 2018. Does AI make PD obsolete? exploring challenges from artificial intelligence to participatory design. In Proceedings of the 15th Participatory Design Conference: Short Papers, Situated Actions, Workshops and Tutorial - Volume 2 (PDC '18). Association for Computing Machinery, New York, NY, USA, Article 8, 1–5. https://doi-org.ezproxy.uio.no/10.1145/3210604.3210646

https://ai.facebook.com/research/no-language-left-behind/

Sebastian Thrun (2004) Toward a Framework for Human-Robot Interaction, Human–Computer Interaction, 19:1-2, 9-24, DOI: 10.1080/07370024.2004.9667338

Mace R. Universal Design, Barrier-Free Environments for Everyone. Los Angeles: Designers West, 1985.

Iwarsson, Susanne, and Agneta Ståhl. "Accessibility, usability and universal design—positioning and definition of concepts describing person-environment relationships." *Disability and rehabilitation* 25.2 (2003): 57-66.

Abascal, & Nicolle, C. (2005). Moving towards inclusive design guidelines for socially and ethically aware HCI. Interacting with Computers : the Interdisciplinary Journal of Human-Computer Interaction, 17(5), 484–505. https://doi.org/10.1016/j.intcom.2005.03.002