IN5620 Individual assignment fall 2022

tinenor

OBS! Jeg er ikke helt ferdig med denne oppgaven, men har forsøkt å skrive noen stikkord på alle oppgaver i alle fall. Bear with me!

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

AI i et historisk perspektiv

Ifølge Jonathan Grudin (2009, s. 2) begynner historien om kunstig intelligens (heretter omtalt som AI) med det potensialet man så i datamaskiner i forbindelse med kodeknekking under andre verdenskrig. Den britiske matematikeren Alan Turing var tidlig ute med å spå at datamaskiner skulle kunne konkurrere med menneskelige intellektet på alle felt. Det var imidlertid amerikaneren John McCarthy som i 1955, sammen med et lite knippe andre akademikere, som først brukte begrepet «artificial intelligence» i innkallelsen til et forskningsprosjekt som gjerne omtales som «the Dorthmouth workshop» (McCarthy et al., 1955). Det tok mange år før Turings spådom skulle bli noe som likner på sann, men de seneste 25 årene har utviklingen innen AI revolusjonerende. Grudin peker på sjakksimulatoren Deep Blue som i 1997 verdensmesteren i sjakk, Gary Kasparov, som et mulig startskudd for en AI-æra som har gitt oss blant annet fjernstyrte rom-roboter, selvkjørende biler og anbefalingsalgoritmer (2009, s. 9).

Å definere AI

Det finnes utallige definisjoner på AI, og jeg vil her se nærmere på tre av dem. La oss først ta en titt på den originale definisjonen til McCarthy et al.: De beskriver det de kaller "the artificial intelligence problem" slik: "[..]how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves" (1955). Denne definisjonen må sees på i lys av at det ennå ikke fantes noen maskin som kunne løse dette problemet, men at McCarthy og hans kolleger hadde et ønske om å få det til. Definisjonen er derfor ikke veldig konkret, men peker på hvilke aspekter ved den menneskelige intelligens de ønsker å utforske med datamaskiner. Allikevel synes jeg at den tar for seg mange relevante egenskaper som også vil kunne kjennetegne mange av dagens AI-løsninger.

Bratteteig og Verne (2018, s.1-2) definerer AI som "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". Det første man kan merke seg er at det har skjedd en signifikant endring i AI beskrevet som et *datasystem* fremfor en *maskin*. Denne definisjonen har allikevel flere fellestrekk med McCarthys. Begge to peker på det menneskelige, samt at maskinen skal løse et problem eller en oppgave. Men i motsetning til McCarthy befinner Bratteteig og Verne seg i en samtid der AI er et faktum. Og der McCarthy kunne sette skillet mellom menneske og maskin basert på at maskinen fortsatt ikke hadde inntatt oppgaver som krever intelligens, må Bratteteig og Verne sette skillet et annet sted. Sammen med deres ståsted som utøvere av det mer menneskesentrerte feltet participatory design er det kanskje derfor ordet *etterlikne* («mimic») blir brukt om maskinens (eller systemets) intelligens.

Eu-kommisjonens AI-ekspertgruppe (High Level Expert Group on AI) foreslo i 2019 følgende definisjon på AI: "Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions" (sette inn kilde). Som de

selv beskriver er denne definisjonen ment å være en utvidet definisjon som skal tydeliggjøre visse aspekter ved AI som forskningsdisiplin, og den bærer også preg av dette. Ekspertgruppa er tydelig opptatt av å beskrive både AIs fysiske form, samt alle mulige måter et AI kan jobbe. I motsetning til de to foregående definisjonene nevnes ikke menneskelig intelligens eller problemløsning i denne definisjonen. Mennesket er kun angitt som designeren av systemet. Det er allikevel interessant at denne definisjonen, i sin detaljrike beskrivelse av hvordan AI fungerer, bruker ord som oppfatte, tolke, resonnere og oppførsel. Dette kan sies å være ord som først og fremst brukes for å beskrive levende vesener, og som til en viss grad bidrar til å personifisere eller levendegjøre datasystemer. Det er også verdt å merke seg hvordan den siste setningen som handler om hvordan AI kan analysere konsekvensene av egne handlinger og basert på dette tilpasse sin oppførsel, kan ses på som en slags moderne speiling av McCarthys mer forenklede beskrivelse av maskiner som kan forbedre seg selv.

Basert på disse tre definisjonen gjør jeg følgende forsøk på å lage min egen definisjon av AI:

AI er et datasystem som, med utgangspunkt i et komplekst mål, samler data og basert på prosessering av denne utfører handlinger for å nå dette målet. Noen former for AI kan også tilpasse sine handlinger basert på konsekvensene av sine tidligere handlinger.

I min definisjon har jeg forsøkt å beholde McCarthys enkelthet, men jeg har valgt å holde mennesket helt utenfor. I stedet har jeg etterstrebet å ta med meg de viktigste aspektene fra ekspertgruppas beskrivelse av hva en datamaskin gjør, samt Bratteteig og Vernes beskrivelse av AI som et datasystem.

Autonomi

Write some sentences about the similarities and differences between human autonomy and machine autonomy. Write some sentences about similarities between the concept automatic and the concept autonomous.

Menneskelig autonomi: å handle uavhengig av andre, bestemme over seg selv.

Maskin-autonomi : å handle uavhengig av mennesket, følge satte regler/møsntre

(bruke Bainbridge og Norman sine artikler om autonomi??)

AI and HCI: two fields divided by a common focus

Artikkelen "AI and HCI: two fields divided by a common focus" av Jonathan Grudin (2009) er først og fremst et historisk perspektiv på hvordan HCI og AI har utviklet seg og sameksistert fra 1950-tallet og frem til artikkelen ble skrevet. Grudin argumenterer for hvordan de to fagfeltene har springer ut fra de samme teoriene, men har endt opp med å konkurrere om økonomiske midler og oppmerksomhet. Særlig peker han på hvordan entusiasmen og optimisme rundt AI har tatt mye oppmerksomhet i perioder, mens HCI har hatt sine glansdager når optimismen har bleknet og AI har vært nede for telling. Grudin skriver at dette har skapt et skille mellom de to fagfeltene som han allikevel gir uttrykk for at kunne ha tjent på å nærme seg hverandre. Avslutningsvis viser han også til tendenser på at dette er i ferd med å skje.

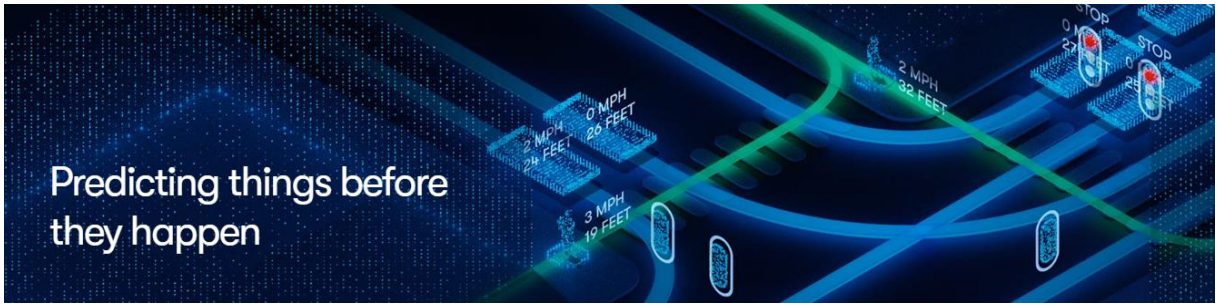
Jeg synes dette var et interessant perspektiv på de to disiplinenes historiske bakgrunn, som jeg ikke allerede var klar over. Ettersom artikkelen er over ti år gammel, blir jeg imidlertid nysgjerrig på hva som har skjedd siden. Den oppsvingen for AI som Grudin beskriver på begynnelsen av 200-tallet må jo kunne sies å ha fortsatt helt frem til i dag. Betyr dette at HCI igjen har kommet i skyggen av sin fascinerende bror eller har de to fagfeltene smeltet mere sammen og dermed latt HCI få være med på oppturen?

Waymo: The world's most experienced driver

Waymo er kanskje et av de mest berømte og vellykkede prosjektene hva gjelder selvkjørende biler. Bedriften startet opp som et Google-prosjekt I 2009, og ble I 2016 et eget teknologiselskap. Siden 2020 har innbyggerne i Phoenix i Arizona kunnet booke en selvkjørende bil som tar dem dit de skal i indre by. Waymo har også begynt å eksperimentere med selvkjørende trailere for frakting av varer.

På Waymos nettsider brukes aldri begrepet AI. De kaller det derimot konsekvent for «autonomous driving technology» har gitt denne et navn; «Waymo Driver». Og Waymo Driver blir på førstesiden omtalt som verdens mest erfarne sjåfør(«The world's most experienced driver»). Dette blir videre utdypet med at det er den samme sjåføren som kjører alle bilene («Same Driver. Different Vehicle.»).

Ettersom Waymo først og fremst tilbyr tjenester, så er det også først og fremst en *tjeneste* Waymo Driver fremstilles som. Og i denne tjenestehistorien er det (naturlig nok) trygghet som står i sentrum. "Informed by unmatched experience and designed with safety at its heart, the Waymo Driver is our autonomous driving technology that never gets drunk, tired, or distracted" star det på nettsiden. Men for å sørge for å underbygge følelsen av trygghet hos både kunder og andre trafikkanter, går de allikevel langt i å beskrive *produktet* Waymo Driver, og forklare hvordan teknologien fungerer. I segmentet som kalles «How it works» endrer den grafiske utformingen av nettsiden seg, og blir mer «teknologisk» med mørk bakgrunn og mengder av prikker som tydelig skal demonstrere mengden av data som ligger bak valgene Waymo Driver tar:



Figur 1: Skjermbilde fra Waymos "How does it work"-segment

Minority report: a 2002 view on AI

Minority report er en film fra 2002, regissert av Steven Spielberg. Handlingen er satt til år 2054, og det amerikanske politiet har gjennom spesialavdeling ved navn Precrime utryddet kriminalitet. Ved å fange synene til tre mennesker med overnaturlige evner, kalt Pre-Cogs, kan Precrime stoppe kriminelle handlinger før de skjer. I løpet av filmen settes det imidlertid spørsmålstegn ved rettsikkerheten i det å arrestere noen som faktisk ikke har utført den kriminelle handlingen ennå, samt hvor sårbare noen blir i et samfunn der noen få har kontroll over et system som styrer over liv og død.

Selv om Precrime baserer seg på evnene til Pre-Cogene, og ikke AI, så er det lett å trekke linjer fra Minority Report og til det at man i noen land har tatt i bruk AI for å styrke forebyggende politiarbeid, ofte kalt «smart policing» (kilde?). Dette har skapt mange etiske diskusjoner som blant annet dreier seg om hvordan AI i noen tilfeller kan føre til at noen sårbare demografiske grupper blir ytterligere stigmatisert og mistenkeliggjort.

I tillegg til denne koblingen spiller AI også mange biroller i Minority Report. Eksempler på dette er blant annet biometrisk identifikasjon (iris-gjenkjenning), spesialtilpassede hologram-reklamer som kan navnet ditt (reklame-algoritmer) og edderkopp-roboter som går inn i utrygge lokasjoner (droner). Da jeg første gang så denne filmen i 2002, opplevde jeg disse som typiske science fiction-elementer, mens de i dag er mer gjenkjennelige enn utrolige.

1.2 Robots and AI systems

Ordet

Fordi ideer om å lage menneskeliknende maskiner med autonomi har eksistert i menneskers fantasi (og dermed i kunsten) mye lenger enn teknologien som kunne virkeliggjære ideen, var det også i kunsten ordet robot først oppsto. Ordet har sin opprinnelse i slaviske språk og på tsjekkisk betyr *robotnik* «arbeider» eller «slave». Det var den tsjekkiske kunstneren Josef Čapek som først brukte begrepet, og broren Karel Čapek

populariserte det i 1920 med sitt teaterstykke *Rossumovi Univerzální Roboti* som handler om en fabrikk som lager menneskekopier kalt «roboti» (Snl, 2022).

Å definere robot

Ettersom begrepet robot først oppsto i kunsten og mer spesifikt I sjangeren science fiction, er det kanskje ikke så rart at det er stor spredning i definisjonene av begrepet. At fungerende roboter bare såvidt var ute av vuggen, er kanskje også grunnen til at the Robot Institute of America i 1979 definerte robot som "An automatic device that performs functions normally ascribed to humans or a machine in the form of a human" (Thrun, 2004, s. 11). På dette tidspunktet var roboter først og fremst maskiner som jobbet på fabrikk og som hadde tatt over komplekse, men repetative oppgaver som tidligere ble utført av mennesker. Definisjonen bærer tydelig preg av dette, i tillegg til at den rommer robotene som folk flest kanskje kjente best; nemlig de fra litteratur og film.

I dag definerer Store norske leksikon en robot som "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. En robot består derfor både av programvare (software) som instruerer den i oppgavene, og maskinvare (hardware) som gir den en «kropp» å bevege" (https://snl.no/robot). Denne definisjonen vitner om den utviklingen som har skjedd siden 1979, og er en mye mer konkret beskrivelse, som også utelukker mange maskiner som kunne falt inn under den forrige definisjonen. Der en oppvaskemaskin kunne vært definert som en automatisk enhet som utfører en oppgave tidligere utført av mennesker, vil den ikke nødvendigvis møte alle kriteriene i SNLs definisjon.

Videre vil jeg definere en robot som *en maskin som kan utføre fysiske oppgaver ved hjelp av sensorer og programvare uten direkte inngripen fra mennesker underveis*. Jeg tenker at en definisjon burde være mer konkret enn 1979-definisjonen, og at vi godt kan holde den menneskeformede maskinen utenfor dette. Jeg synes definisjonen til SNL gir mye mening, samtidig savner jeg en understreking av at oppgaven(e) som roboten utfører er fysisk(e). Jeg har også valgt å spesifisere at en robot utfører oppgaver uten menneskelig inngripen underveis for å utelukke fjernstyrte enheter, som jeg tenker at er et separat konsept.

AI vs. Robot

Discuss the relation between AI and Robots. Is "a robot" different from "an AI"? In what ways are they different and similar? Bring in the definitions that you described earlier about robots and AI for this discussion.

Forskjeller: Den fysiske formen. En robot må ha en fysisk form, en AI defineres ikke av det. En robot kan ha AI, men AI trenger ikke være en robot.

Likheter: Autonomi. De utfører begge oppgaver. Og de skal kunne gjøre det uten at mennesker må bidra underveis.

AX1: The next generation of ultra-high precision weeding

Find one contemporary physical robot, either described in a research article – or a commercial robot, and describe how this robot moves and how a human user is interacting and using the robot in a specific situation.



AX1 – norsk landbruksrobot. Bruker AI-teknologi og bildegjenkjenning til å sprøyte små doser ugressmidler på ugresset og ikke på avlingen.

Interaksjon: Poenget er at bonden ikke skal trenge å interagere så mye med roboten mens den jobber, slik at hen kan bruke tiden på noe annet. Jeg antar derfor at interaksjonen skjer i forbindelse med igangsettelse og avslutning av en økt, og eventuelt i en foreberedelsesfase, der man lager avgrensninger for hvor roboten skal jobbe (https://www.kiltersystems.com/ax1) .

1.3 Universal Design and AI systems

Å definere Universell Utforming

§ 17 i lov om likestilling og forbud mot diskriminering: "Med universell utforming menes utforming eller tilrettelegging av hovedløsningen i de fysiske forholdene, inkludert informasjons- og kommunikasjonsteknologi (IKT), slik at virksomhetens alminnelige funksjoner kan benyttes av flest mulig, uavhengig av funksjonsnedsettelse" (https://www.regjeringen.no/no/dokumenter/veileder-i-universell-utforming/id2850026/?ch=3). Jeg forstår denne definisjonen som at UU handler om at fysiske og digitale miljøer skal utformes på en slik måte at en størst mulig del av befolkningen skal kunne bruke dem, uansett hva slags fysiske eller psykiske forutsetninger de måtte ha. Klarer man å oppnå dette, vil man demokratisere disse miljøene og ikke ekskludere noen.

AI som inkluderer

Ettersom AI I stor grad handler om mønstergjenkjenning er det mange som ser potensiale i å bruke AI til å kjenne igjen både omgivelser, bevegelse og til og med menneskelige følelser, for så å gjenskape disse. For mennesker som har nedsatt syn, hørsel, bevegelse eller andre kognitive evner, kan teknologien da tilby muligheten til å støtte dem i hverdagen. Eksempler på dette er kameraer som ved hjelp av bildegjenkjenning kan fortelle hva som er foran linsa (https://www.microsoft.com/en-us/ai/seeing-ai) eller proteser med innebygget AI for å forbedre brukerens kontroll over bevegelsene (https://www.discovermagazine.com/technology/ai-is-revolutionizing-prosthetic-arm-control).

Eksemplene over kan, hvis de fungerer godt nok, hjelpe til med å inkludere flere i større deler av samfunnet vårt. Samtidig finnes det også eksempler på bruk av AI som har vist seg å være ekskluderende. Et eksempel på dette er tidligere det tidligere nevnte fenomenet «smart policing» der AI bistår politiet med å forutse hvilke områder det er mest sannsynlig at kriminalitet vil oppstå basert på analyse av data. Men det har blitt pekt på at dersom politiet legger ekstra ressurser på å være til stede i visse områder, så vil de også finne mer kriminalitet der. Dette kan føre til at befolkningen i disse områdene (som gjerne i utgangspunktet er sårbare og til en viss grad ekskludert fra deler av samfunnets gode selskap) risikerer å bli stigmatisert og dermed ytterligere ekskludert. Et annet eksempel er når en AI-løsning som bruker optiske sensorer ikke fungerer likt for folk med ulik hudfarge (https://gizmodo.com/why-cant-this-soap-dispenser-identify-dark-skin-1797931773). Dersom noen grupper med mennesker blir ikke kan bruke såpe på lik linje med sine kolleger eller medborgere, vil AI ha bidratt til å eksludere.

Understanding “understand”

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?

Oppfattbart, operativt, og forståelig. Forstå er et steg opp fra å oppfatte. Kan inkludere mye mer enn å forstå en løsning, men kan innebære å forstå et system eller viktige deler av samfunnet.

1.4 Guideline for Human-AI interaction

Jeg har valgt guideline nr 11: “Make clear why the system did what it did”, som er under kategorien “when wrong”. Jeg tenker at det er veldig viktig som designer å ha tenkt nøye igjennom hva som skal skje når et system feiler. Videre tenker jeg at særlig i arbeidet med AI er det viktig å vise en grad av transparens når systemet ikke oppnår det det skal. Dette vil kunne dempe den følelse av fremmedgjøring som AI i sine verste former kan skape.

Eksempel

Search, and find one set of HCI design guidelines. Discuss briefly similarities and differences between the HCI design guidelines and the Human-AI interaction guidelines.

Bruke Don Normans 7 designprinsipper?

Referanser

(OBS! Må ryddes opp i!)

Bratteteig, T. & Verne, G. (2018). Does AI make PD obsolete? Exploring challenges from Artificial Intelligence to Participatory Design.

Grudin, J. (2009). AI and HCI: two fields divided by a common focus.

High-Level Expert Group on Artificial Intelligence https://digital-strategy.ec.europa.eu/en/library/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines?fbclid=IwAR3oPKZL2dYvaN89dkNwbelVZLWyXOB7qtefz309Gc0IWmbLxac1ONJJHqk

Kilter https://www.kiltersystems.com/ax1

Lovdata

McCarthy, J., Minsky, M., Rochester, N. & Shannon, C. (1955). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence.

Store Norske Leksikon

Thrun, S. (2004). Toward a Framework for Human–Robot Interaction

Waymo https://waymo.com/