Iterasjon 3

Modul 1

Konsepter, definisjoner og historie innfor interakasjon med AI og autonome systemer

AI- definisjonshistorie

Alan Turing publiserte en artikkel i 1950 der han stilte spørsmål rundt hvor vidt det var mulig å lage en maskin som tenkte selv, men begrepet kunstig intelligens (AI) ble først introdusert i 1956 av logiker og matematiker John McCarthy (Cordeschi, 2007). Introduksjonen innebar optimistiske idéer og forestillinger AI aldri klarte å måle seg med, og historien har dermed vært preget av skuffelse og bølgedaler. Teknologien har enda langt igjen for å kunne måle seg med de optimistiske forventningene, men har samtidig kommet svært langt.

AI- 4 definisjoner

John McCarthy definerte i 2004 AI som "vitenskapen og konstruksjonen bak å lage intelligente maskiner, spesielt intelligente dataprogrammer (Sabanovic et al., 2012). Det er relatert til den lignende oppgaven med å bruke datamaskiner for å forstå menneskelig intelligens, men AI trenger ikke å begrense seg til metoder som er biologisk observerbare." McCarthy har siden han introduserte begrepet i 1956 vært viktig for utviklingen av AI, og definisjonen bør anerkjennes. Tar man dog AI sin utvikling siden 2004 i betraktning kan det være nødvendig å kaste nytt lys på definisjonen.

Marvin Lee Minsky var en informatiker med særlig interesse for AI, og arbeidet ved MIT-fakultetet fra 1958 til sin død (Seising, 2016). Han startet i 1959 MIT Computer Science and Artificial Intelligence Labratory sammen med John McCarthy. I 1968 definerte han kunstig intelligens som "vitenskapen ved å få maskiner til å gjøre ting som ville kreve intelligens hvis de ble gjort av menn". Definisjonen deler likheter med McCarthy sin, og det kommer frem at de har jobbet tett sammen med fagfeltet.

Andreas Kaplan og Michael Haenlein, professorer med fokus på AI, definerte i 2018 AI som "et systems evne til å korrekt tolke eksterne data, å lære av slike data, og å bruke denne kunnskapen til å oppnå spesifikke mål og oppgaver gjennom fleksibel tilpasning" (Haenlein & Kaplan, 2018). Definisjonen skiller seg fra de to andre definisjonene ved at den er mer spesifikk og tilpasset vår samtidens teknologi.

Feltets tverrfaglighet, bredde og hyppige utvikling gjør det utfordrende å stadfeste en definisjon. Elon Musk, en moderne pioner innenfor AI, har utfordringer med å definere AI, hvilket forsterker mitt inntrykk av hvor flytende begrepet er. Ved å bygge definisjonen min på en kombinasjon av samtiden og de nevnte definisjonene, vil jeg likevel definere AI som maskiner og teknologi som gjennom aktiv læring og analysering av store mengder data fleksibelt tilpasser seg til å løse behov på et vis som fremstår så intelligent at det kunne vært gjort av et menneske.

AI- autonomi og automatikk

Med menneskelig autonomi menes individets evne til å være selvstyrende og moralsk uavhengig. Dersom man er autonom bestemmer man over seg selv og sine egne handlinger. En maskin anses derimot som autonom når den har evnen til å tilegne seg nok kunnskap om miljøet sitt til å være i stand til å gjennomføre oppgavene den er satt til å gjøre, selv når den nødvendige kunnskapen ikke var der fra start.

Til likhet med autonomi, fungerer noe som er automatisk helt av seg selv. Forskjellen ligger i mengde tilpasning, læring og beslutningstaking som er integrert i systemet. Autonome systemer har færre

begrensninger enn automatiske. De er dynamiske, lærer fort, og kan gjerne begi seg forbi sitt opprinnelige felt. Automatiske systemer begrenses av satte rammer og oppgaver, og valgene og oppgavene den gjennomfører er mer basert på forhåndsbestemt heuristikk.

Omtale av artikkel

I "Some Brief Thoughts on the Past and Future of Human-Robot Interaction" fra 2018 tar Kerstin Dautenhahn tanker og meninger om utviklingen av HRI hittil og veien videre. Hun adresserer blant annet suksessfaktorer, utfordringer, kjernebegreper og standarder. Hun stempler fagfeltet som en suksessfaktor de siste årene, og peker på årlige konferanser der fagfelt deler kunnskap, erfaringer og bekymringer på tvers av hverandre. Ironisk nok anses denne tverrfagligheten som en hovedutfordring, siden de ulike fagfeltenes forskjellige metoder og fokusområder gjør det vanskelig å felle fellesvurderinger over HRI. Tverrfagligheten har samtidig medført utfordringer tilknyttet begrepsstadfesting, for sentrale begreper som "robot". At eksperimentell psykologi er å anse som en gullstandard i HRI vekker bekymringer, siden det ikke tar høyde for virkelig interaksjoner mellom virkelige mennesker og virkelig komplekse roboter i virkelige omgivelser. Kvalitativ forskning med fokus på virkelige interaksjoner trekkes frem som nødvendig, til tross for at det er mer ressurskrevende. Til tross for utfordringene tverrfagligheten bringer, anser Dautenhahn dette som positivt for HRI, og hun understreker viktigheten av åpent sinn i møte med tilnærminger fra andre fagfelt.

Jeg har utfordringer med å forstå hvordan andre fagfelt er så verdifulle som det gis inntrykk av, og savner mer konkrete eksempler på informasjonsutveksling av verdi. Det kan være krevende å visualisere hvordan fagfolk uten teknologisk bakgrunn kan bidra med utviklingen av HRI. Med robot som flytende begrep savner jeg dessuten en klarere forståelse av hva som menes med robot og dermed også interaksjoner innen HRI. Dautenhahn nevner at den kvalitative forskningen bør bruke komplekse, intelligente og multitaskende systemer. Dette er forståelig nok, men når interaksjonen skal skje "face 2 face" maler hun et urealistisk og futuristisk bilde av en sci-fi robot som lever i tråd med fortidens forventninger til AI. Å forestille seg at denne interaksjonen skal skje i naturlige omgivelser blir dermed vanskelig. Ser jeg forbi dette er jeg ikke uenig med Dautenhahn.

AIBrain - moderne AI selskap

AIBrain ble stiftet i 2012, og har i mange år vært anerkjent som et av de beste AI-selskapene. Selskaper presenterer AI som et produkt som skal øke brukernes intelligens langs fire dimensjoner - kognitivt, sosialt, fysisk og emosjonelt. De viser til nevrale og kognitive vitenskaper for å støtte påstanden deres. Åpenhet og dokumentasjonen av teknologiens virkemåte gjør at selskapet fremstår profesjonelt og ærlig, istedenfor å presentere AI som "mystisk teknologi" der virkemåten er skjult bak kulissene. Landingssiden deres portretterer et bilde av en robot-hånd som speiler en menneskehånd på andre siden av et futuristisk grensesnitt. Bildet er åpent for tolkning, men representerer på mange måter den klassiske forståelsen av AI som menneskets erstatter. Utover dette presenteres AI på en profesjonell og realistisk måte.

I, Robot (2004) - film om samspill mellom AI og mennesker

Filmen finner sted i 2035 der menneskelignende roboter lever side om side med mennesker i samfunnet. Robotene fungerer er svært avanserte, og fungerer som brede og spesifikke oppgaveløsere. Samspillet mellom mennesker og roboter er tilsynelatende naturlig, og robotene fremstår som vennlige hjelpere uten følelser. Interaksjonen dem i mellom skjer gjennom tale og kroppspråk, men robotene virker ikke avhengig av denne kommunikasjonen for å forstå hva de skal gjøre. De fremstår svært autonome i form av at de tar del i samfunnet og løser oppgaver på naturlig og logisk vis, slik et menneske ville gjort, men innenfor gitte etiske rammer. Filmen tar et vendepunkt

når en teknofobisk etterforsker tar en robot i å begå et drap, og samspillet endres. Tilliten til robotene er svekket, og maktbalansen er snudd. Vendepunktet beskriver AI som har utviklet seg i mot sin hensikt og vendt seg mot mennesket, der roboten har beveget seg utenfor sine gitte rammer og handler ut ifra "fri vilje".

Robot - definisjonens oppstandelse

Ordet robot ble først introdusert for publikum av den tsjekkiske forfatteren Karel Capek i skuespillet R.U.R i 1920 (Moran, 2007). Ordet stammer fra det slaviske ordet robota, som betyr "arbeid", eller "tvangsarbeid". Stykket viser en fabrikk som lager kunstige mennesker som ligner en premoderne forestilling av robot, slik man ser i filmen I, Robot.

Robot - 3 definisjoner

Dautenhahn anerkjenner at man innenfor HRI ikke har en felles definisjon for robot. Dette medfører at et bredt spekter av variasjoner på tvers av fagfelt og eksperter. Den offisielle definisjonen er imidlertid "Actuated mechanism programmable in two or more aces with a deegree of autonomy, moving within its environment, to perform indented tasks." Definisjonen beskriver et maskineri som kan programmeres til å gjennomføre tiltenkt oppgave innenfor sitt miljø med en grad av autonomi. SNL definerer robot som "Robot er en maskin som kan utføre handlinger automatisk, og som ved hjelp av elektroniske «sanser» og «oppskrifter» på hva de skal gjøre (algoritmer) kan bearbeide inntrykk fra omgivelsene, som igjen påvirker hvordan disse oppgavene utføres". Her beskrives robot som mer automatisk enn autonomt, men til likhet med den offisielle definisjonen skal roboten være i stand til å dynamisk gjennomføre oppgaver basert på omgivelsene sine. Basert på de nevnte definisjonene vil jeg definere en robot som en autonom maskin som er i stand til å sanse omgivelsene sine, utføre beregninger for å ta beslutninger og utføre fysiske handlinger i den virkelige verden.

Relasjon mellom AI og roboter

AI og robotikk eksisterer både uavhengig og avhengig av hverandre. Ved å kombinere AI med robotikk får man en kunstig intelligent robot med et høyt nivå av autonomi som vil være i stand til å optimalisere oppgaver den er satt til å gjøre og "lære" de beste metodene. I kombinasjon vil AI vil på sett og vis fungere som robotens hjerne, og sensorene og de mekaniske delene som kroppen. Med en lite fastsatt definisjon av robot, og en flytende forståelse av AI er det imidlertid utfordrende å sette grensen mellom hvor grensen mellom robot og AI går. Forskjellen blir dog tydelig ved at AI ikke interagerer med fysiske miljøer, f.eks gjennom algoritmer i dataprogrammer, samtidig som roboter gjennomfører fysiske oppgaver. Tar man min egen, samt den offisielle definisjonen av robot i betraktning, kommer det frem at fysiske oppgaver må fremkomme for å kunne klassifisere noe som robot. Definisjonene betrakter samtidig roboter som autonome, til likhet med definisjonene av AI.

Roxo the FedEx SameDay bot - moderne fysisk robot

Roxo er FedEx sitt forslag på futuristisk logistikk. Roboten er designet for å levere pakker i en radius på tre til fem kilometere fra en forhandlers beliggenhet innen et døgn. Roboten skal kunne kjøre på fortau, sykkelfelt og veikanter, og beveger seg som en liten på fire hjul. Den har to støttehjul for å kunne manøvrere over objekter som fortauskanter, og sakker farten før en endring i bevegelse, slik at bevegelsene dens gir et visuelt naturlig inntrykk. Roboten kommuniserer gjennom en stor og tydelig skjerm, samtidig som den har en begrenset talefunksjon.

UU - definisjoner

Universell utforming (UU) er nødvendig for inkluderende design i flere fagfelt, slik som IKT og byggebransjen. Forskriften om UU innen KTI ble oppdatert i 2022, og definerer UU som "Utforming eller tilrettelegging av hovedløsningen i informasjons- og kommunikasjonsteknologi slik at virksomhetens alminnelige funksjon kan benyttes av flest mulig". Definisjonen vektlegger at flest mulig, altså også de med bruksvansker slik som kognitive utfordringer, skal kunne benytte seg av virksomhetens digitale hovedløsning. Definisjonen deler mye med definisjonen i diskriminerings- og tilgjengelighetsloven som ble revidert i 2013, der den tydeligste forskjellen er at denne definisjonen også inkluderer fysiske forhold: «Med universell utforming menes utforming eller tilrettelegging av hovedløsningen i de fysiske forholdene, herunder informasjons- og kommunikasjonsteknologi (IKT), slik at virksomhetens alminnelige funksjon kan benyttes av flest mulig». Ronald Lawrence Mace var en av grunnleggerne bak UU, og definerte konseptet som "design som kan brukes av alle mennesker, i størst mulig grad, uten behov for tilpasning eller spesialisert design". Til tross for at definisjonen er over 25 år gammel, er den fremdeles relativt presis og dagsaktuell, og det viser hvilken felles forståelse man har av UU både på tvers av og innad i fagfelt. I forhold til inkludering handler UU om å designe for at flest mulig skal kunne bruke en tjeneste, til tross for at de har ulike forutsetninger.

AI- menneskelighet og potensiale for inkludering

Siden AI ble introdusert har offentligheten hatt optimistiske forventninger til utviklingen. AI har kommet langt, men klarer enda ikke gjenspeile mennesker på en naturlig måte. Vår evne til persepsjon, bevegelse, og følelser er faktorer som skiller mennesker fra dyr - og teknologi. Persepsjon er sanseinntrykk og den påfølgende tolkningen av disse, og skjer i to trinn: gjennom stimuli av ett eller flere sanseorgan, og at stimuleringen tolkes og resulterer i en opplevelse. Vi har ulike forutsetninger for persepsjon, basert på tidligere inntrykk, og tilgjengelige sanser. En AI, eller en robot, vil også ha ulike forutsetninger for persepsjon. Denne vil da baseres på hvilke komponenter den bruker til å sanse og analysere omgivelsene sine, og hva den har lært. Dagens teknologi gjør at AI har potensiale for enda flere og sterkere sanser enn oss. AI bruker dyp læring for å tolke mønstre i data for å ta en endelig avgjørelse, hvilket på mange måter ligner menneskelig beslutningspsykologi. AI bruker samtidig innkodete standardinstruksjoner for bearbeiding av logikk og begrunnelse, hvilket ligner menneskelig rasjonell intelligens. Det kan virke som AI sine forutsetninger for å oppnå menneskelig persepsjon er gode, om ikke bedre enn våre egne, men at tolkningen av sanseinntrykkene deres ikke resulterer i en opplevelse er avgjørende for hvorfor de likevel ikke er i stand til det. Det finnes ingen AI som er i stand til å bevege seg slik som mennesket gjør. At bevegelsene våre er komplekse gjennom konstant kalibrering av balanse, fart, mimikk og kroppspråk gjør det vanskelig å kopiere bevegelsene våre. Det finnes menneskelignende roboter, slik som Ameca the humanoid, som klarer å etterligne menneskelig atferd, men som røpes av unaturlig ansiktsmimikk, og en manglende helhetlig kroppsbevegelse. Roboten vil heller ikke være i stand til så mye mer enn å kopiere atferd og å gjennomføre simple oppgaver, så de fysiske begrensningene hindrer den fra i å f.eks spille fotball. AI har ingen evne til å føle, kun evnen til å vite hvordan den skal respondere til sin egen tolkning av dataen den registrerer. Våre følelser er basert på hormoner og er ikke noe vi nødvendigvis har kontroll over, til motsetning fra AI sin frie tolkning av binære tall. Dersom man vender blikket mot vanligere KI, som ikke er skapt for å fysisk etterligne et menneske, slik som en smarthøyttaler, blir den manglende tilstedeværelsen av de menneskelige egenskapene enda tydeligere. Til tross for manglende menneskelige egenskaper, kan AI ha potensiale for inkludering. Et godt eksempel er gjennom appen "seeing AI", som er til for å assistere blinde. Appen har flere moduser, og forteller hva som oppfattes foran skjermen, slik at den blinde vet hva den står ovenfor. Smarttelefoner med talegjenkjenning og bildegjenkjenning forenkler også hverdagen til utallige med nedsatt funksjonsevne gjennom AI.

AI- inkluderende eller ekskluderende?

AI har både potensiale for inkludering og ekskludering. Funksjoner som talegjenkjenning og bildegjenkjenning for blinde er tydelige eksempler på hvordan AI kan inkludere de med særskilte behov. Dette er imidlertid AI som er designet for spesifikk inkludering. Ved design for massene er AI svært mottagelig for å ekskludere siden den lærer basert på data fra massene. Ofte brukes normalfordeling i behandlingen av datasettene, og dette fører til at de som avviker fra normalbruk blir utelatt. Samtidig har AI potensiale for rasisme og sexisme. Dette kan både være basert på datasett AI-en selv har plukket opp som viser korrelasjon mellom kjønn eller rase og verdien AI-en er interessert i, eller at feil i algoritmene får AI-en til å diskriminere basert på rammene den er gitt. Eksempel på rasisme integrert i AI ble i 2019 funnet i det amerikanske helsevesenet på et system brukt på over 200 millioner mennesker for å forutsi hvilke pasienter som hadde behov for medisinsk behandling. Algoritmen prioriterte hvite pasienter fremfor svart pasienter basert på helsekostnadshistorie som viste seg å som regel være høyere blant hvite mennesker. Selv om rase ikke var en variabel i algoritmen prioriterte algoritmen de med høyest helsekostnadshistorie, som altså viste seg å korrelere med rase. Det ble i 2015 avduket feil i Amazon sitt ansettelsessystem som til likhet skilte mellom menn og kvinner basert på antall søknader som var innsendt de siste 10 årene (Wonchul & Keeheon, 2019). Systemet var opptrent basert på søknadene som var sendt inn det siste tiåret, og siden majoriteten av søknadene var fra menn var systemet opptrent til å favorisere menn.

AI- kan maskiner forstå?

Forståelighet er som en av fire grunnprinsipper en grunnpilar for tilgjengelighet på nett. Prinsipper dreier seg i all hovedsak om at brukere må være i stand til å forstå informasjonen så vel som hvordan grensesnittet opereres. Forståelse på begge ender av grensesnittet er viktig, og setter skillet mellom å forstå og forståelse. Brukeren må forstå både informasjonen som gis og grensesnittet som tilbys, men tjenesten må også være i stand til å forstå brukeren, også dersom brukeren velger å gi input gjennom utradisjonelle metoder. Hvor vidt maskiner har kapasitet for forståelse kan diskuteres og er i stor grad et definisjonsspørsmål. Hva er en maskin og hva er forståelse? Maskiner kommer i alle størrelser og varianter, og det gjør også forståelse. To individer kan ha samme opplevelse uten å dele den samme forståelsen av opplevelsen. Man har enten forståelse for eller av noe. Om man har forståelse for noe har man empati, og forståelse for personer, følelser eller folks situasjoner, og gjerne ved å tidligere ha vært i lignende situasjon selv. I dagligtale brukes uttrykk som å "sette seg i andres sko eller fotspor", og dette er noe en maskin vil ha utfordringer med å gjøre, siden den naturligvis ikke har levd et liv på linje med oss selv. Ved å ha forståelse av noe har man forståelse av kunnskap og oppfatninger, slik som grammatikk eller geografi. En maskin har absolutt kapasitet for kunnskap, men hvor vidt den faktisk forstår kunnskapen sår jeg tvil over.

Retningslinjer for menneske-AI interaksjon

Prinsipp 13, "learn from user behaviour", i Microsoft sine retningslinjer for menneske-AI interaksjon låter "Personalize the users experience by learning from their actions over time", og dreier seg om å gi brukeren skreddersydde opplevelser basert på tidligere bruk. Brukerens handlinger under bruk gjenspeiler ikke nødvendigvis brukerens behov og ønsker. Bruken kan ha vært basert på at brukeren var nysgjerrig, distrahert, klikket feil, eller var i en annen uvanlig situasjon som gjorde at bruken skilte seg fra normalbruken. Brukeren kan dermed skape sin egen bias basert på et usedvanlig bruksmønster som systemet ikke klarer å skille fra normal bruk. Et typisk eksempel på dette er når man får opp haugevis av reklame på nett for ting man har klikket på ved uhell, eller kun klikket på av nysgjerrighet. Slik atferd redegjør ikke prinsippets beskrivelse for, så et logisk alternativt eksempel kunne vært "Personalized user experience based on user preferences and repetitive actions over time".

Shneidermans åtte retningslinjer innenfor HCI er skapt for å hjelpe designere med å løse problemer, og har flere likheter og forskjeller med Microsoft sine menneske-AI interaksjonsretningslinjer. Begge settene med retningslinjer setter brukbarhet og brukerens forståelse i sentrum. Eksempelvis understreker begge settene at informativ feedback er nødvendig for at brukeren skal forstå hva som skjer bak skjermen, eller at man enkelt skal kunne avslutte en prosess. Til forskjell fra Microsoft sine retningslinjer fokuserer HCI-retningslinjene på consistency. Det er samtidig utfordrende for AI å være consistent når teknologien bygger på stadig utvikling. Microsoft sin imøtekommelse med utfordringen adresseres dog av prinsipp 14, som foreslår at oppdateringer og tilpasninger må skje på varsomt vis, samtidig som settet med retningslinjer understreker viktigheten av å unngå å forvirre brukeren ved å alltid informere om hvorfor systemer oppfører seg som det gjør.

Modul 2

Endringer siden iterasjon 1

Basert på tilbakemeldingen fra iterasjon 1 har jeg skrevet modul 2 med et høyere fokus på kildeføring underveis, slik at det er tydeligere hvor informasjonen kommer fra. Jeg har også fylt inn noen manglende kilder i modul 1, og gått over til å bruke «AI» istedenfor «KI».

Karakteristikker ved AI-infunderte systemer

AI-infunderte systemer defineres som systemer som utnytter AI-egenskaper som direkte eksponeres for sluttbrukeren (Amershi et al., 2019). Systemene har 4 tydelige kjennetegn; de lærer, forbedrer seg, benytter "black box"-prosesser, og er drevet av store datasett (Kocielnik et al., 2019). Systemene er avhengig av bruk over tid for å lære. Dersom systemet ikke brukes, vil det ikke få ny treningsdata, og ikke ha muligheten til å forbedre seg. AI-infunderte systemer er designet for dynamikk, og har kapasitet til å utvikle seg over tid. Målet med utviklingen er at systemet skal forbedre seg basert på historisk bruk, men det skjer også at systemet blir dårligere, og tilegner seg uønskede egenskaper. For at systemet skal forverres må uønsket bruk og input forekomme, slik som når Twitter-bots har blitt matet med sexistisk og rasistisk input (Følstad & Brandtzaeg, 2017). Slik bruk må forekomme gjentatte ganger for at systemet skal påvirkes, ettersom det gradvis utvikler seg til å bedre tilpasse seg brukernes bruk. At systemet er designet for å forbedre seg gjør at feil er uunngåelig. Systemene benytter seg av black box-prosesser som gjør det umulig å vite nøyaktig hvorfor systemet oppfører seg som det gjør. Alt systemet gjør loggføres, men man har ikke innsyn i hvordan algoritmene påvirkes av opplæringen. For at systemene faktisk skal fungere avhenger det av store datasett, som typisk hentes fra nett, gjennom større brukergrupper, eller som direkte mates inn i systemet for å trene det opp. I disse datasettene ligger systemets kunnskap om omverdenen, og datasettenes størrelse blir dermed avgjørende for å maksimere kunnskapens nøyaktighet. Dette er grunnen til at systemer med mindre datasett er lettpåvirkelige for innmating av uønsket data.

TikTok som AI-infundert system

Videoplattformen TikTok eksponerer brukerne sine direkte for AI-egenskaper, og følger dermed Amershi (2019) sin definisjon av et AI infundert system. Plattformen serverer en skreddersydd feed, kalt "for you page" til brukeren, og legger ikke skjul på hvorfor den viser videoene den gjør, men gir en forklaring på at disse videoene vises siden brukeren har interagert med lignende innhold. Systemet oppfører seg i henhold til Kolcielnik sine 4 nøkkelegenskaper (2019), hvilket påvirker brukerne i mange retninger. Systemet kategoriserer innhold basert på hashtags, kommentarer, skapere og videoenes lyd. Brukernes interaksjoner utgjør videre hva TikTok velger å gjøre med disse videone, og dette er bakteppet for det store og tilsynelatende tilfeldige spriket i antall visninger på relativt like videoer. Interaksjonen kan skje på flere måter, og forteller TikTok om hva brukeren foretrekker å se av innhold. Interaksjon kan skje direkte ved at brukeren liker, kommentarer, deler,

lagrer, eller interagerer med innholdets skaper, eller inderekte ved at brukeren har videoen på repeat. Slik interaksjon vil fortelle TikTok at det foreslåtte innholdet var korrekt, og TikTok vil fortsette å foreslå lignende innhold. Brukeren kan også fortelle brukeren at det foreslåtte innholdet var feil, gjennom å rapportere, trykke ikke interessert, eller ved å ikke se videoen og bla videre, og TikTok vil derette foreslå nytt innhold frem til det får feedback om at innholdet er relevant. TikTok er blant verdens mest populære apper, og har med over 1 milliard daglige aktive brukere (Ruby, 2022) enorme datasett.

Basert på brukerens interaksjoner vil ikke TikTok kun tilpasse den ene brukerens feed, men også andre brukere som interagerer med lignende innhold. Dette kan ha både positive og negative konsekvenser. Innholdet vil som regel ha større treffsikkerhet, siden det allerede har blitt "validert" av andre brukere. Utfordringen ligger imidlertid i at brukere ikke kun interagerer med videoer basert på faktisk interesse. Videoer som oppleves som sjokkerende, nysgjerrighetsvekkende, dumme, eller unormale vil også fange brukeres interesse. Med dette har TikTok høy kapasitet for å bli en plattform som ikke kun viser morsomt og lærerikt innhold, men også promoterer mobbing, rasisme, sexisme, grafisk innhold, ekstremisme, og lignende. Når brukere stadig eksponeres for slikt innhold kan man risikere at de, til likhet med AI-infunderte systemer, gradvis påvirkes til å absorbere uønskede holdninger. Selv uten uønsket innhold kan TikTok og systemets evne til å tilpasse seg brukere ha negative konsekvenser i den åpenbare form av at det kan være avhengighetsskapende. Til forskjell fra mange andre sosiale medier er TikTok svært direkte og frempå, og vil forsøke å fange brukerens interesse med en kort video fra sekundet appen åpnes. TikTok kan anses som en dopamin-fabrikk, der hjernen stadig belønnes for å ta til seg ny informasjon og innfri et underholdningsbehov (Qin et al., 2022). Dopaminlageret vårt er ikke evig, og vil gå tomt. Når dette skjer søker vi mer, og slik starter typisk en avhengighet. Med en endeløs rekke av korte videoer som over bruk øker i treffsikkerhet er det høy risiko for at man bruker mer tid enn planlagt i appen, og man kan stille seg det åpenbare spørsmålet om hvordan navnet "TikTok" korresponderer med tidsbruk.

Hovedpoeng fra Amersi og Kocielnik sine artikler

Amersi (Amersi et al., 2019) adresserer muligheter og utfordringer ved design for interaksjon med AI, og foreslår i lys av dette 18 retningslinjer. Retningslinjene er generell forslag forfatterne håper kan fungere som en ressurs for aktørene som er involvert i design av menneske-AI interaksjon. Retningslinjene er baset på over 20 år med forskning innenfor HCI og menneske-AI interaksjon. De er opprettet i 4 faser med flere iterasjoner og runder med iterering, hvilket gjør dem valide. Behovet for retningslinjer bunnet i at interaksjonen brøt med eksisterende retningslinjer innenfor usability, slik som consistency, siden systemet stadig var under endring og tilpasset seg ulike brukere. Til tross for at forfatterne ønsker at retningslinjene tas i bruk anerkjenner de at nye retningslinjer kan være nødvendig for AI-baserte systemer med høyere risiko og behov for spesialisering, slik som operasjonsutstyr, autonome kjøretøy, og finanssystemer.

Kocielnik (Kocielnik et al., 2019) utforsker hvordan man kan forberede brukere av AI-infunderte systemer på feil, og sentrerer spørsmålet rundt en planleggingassistent som detekterer og oppretter planer basert på innhold i mailen. Han tilnærmer seg problemstillingen med to utgaver av systemet der de har samme oppgave, men ulik fremgangsmåte. Begge systemene har som formål å detektere og opprette planer i kalenderen basert på innhold i mailen, men det ene systemet er åpent og passivt, og det andre mer lukket og aggressivt. Det åpne systemet er mer gjennomsiktig, og bruker en nøyaktighetsindikator for å dele muligheten for at det tar feil med brukeren, men det lukkede systemet gjør ikke dette. På denne måten utforsker forfatterne hvordan man bedre kan skape realistiske forventninger for brukere, og forberede dem på at systemet ikke er feilfritt. Det åpne systemet deler samtidig hva forslagene er basert på og tillater brukeren å justere hvor aktivt systemet skal oppføre seg, til motsetning fra det lukkede systemet.

TikTok sin tilnærming til retningslinjene
Jeg har valgt ut en retningslinje jeg mener TikTok følger, og har godt av å følge, og en retningslinje jeg tror TikTok ville hatt godt av å følge.

G13 – Learn from user behaviour. Personalize the user's experience by lerning from their actions over time.

- Dette er en retningslinje jeg mener TikTok følger, og har tjent godt av. TikTok sin suksess sentrerer seg rundt at hver bruker har sine egne skreddersydde feeds med innhold tilpasset historikk fra tidligere interaksjoner og viralt innhold. Med mindre skreddersydd innhold ville brukeren mistet interesse, og redusert bruken, hvilket kunne ført til reduserte datasett og en reduksjon i systemets evne til å forbedre seg.

G17 – provide global controls. Allow the user to globally customize what the AI system monitors and how it behaves.

- Dette er en retningslinje jeg mener TikTok ikke følger, og ville ha godt av å følge i form av en filtreringsfunksjon for uønsket innhold. Slik kunne TikTok vernet sensitive og yngre brukere fra uønsket innhold. Det ville samtidig fjernet uønsket interaksjon med videoer som ikke er av interesse, og redusert skjevdata som bidrar til å popularisere slikt innhold.

Bender et al sin argumentasjon
Bender et al (2021) kaster et kritisk blikk på AI infunderte systemer som baserer seg på store språkmodeller, og identifiserer flere problemområder. Hovedutfordringen med veksten av språkmodellene er at de er lite miljøvennlige og kostbare. Kostnadene gjør inngangsterskelen høyere for nye aktører, og begrenser hvem som kan bidra med å utvikle feltet videre. De foreslår rapportering av kostnader og en ressursbasert selvevaluering som en potensiell løsning. Datasettenes størrelse gjør det samtidig utfordrende å få innsikt og oversikt over systemenes treningsdata. Istedenfor å ta absorbere all tilgjengelig data, foreslår forfatterne å investere ressurser i å nøye dokumentere datasett. Forfatterne er bekymret for systemenes evne til å tilegne seg rasistiske, sexistiske og ekstremistiske verdenssyn, og håper at bidraget deres kan bidra med å forbedre dette.

Sentrale utfordringer i utformingen av chatbots
Å designe for sømløs og naturlig interaksjon med chatbots er utfordrende av flere grunner. En av nøkkelutfordringene er at teknologien er relativt ny og skiller seg fra tradisjonelt design i HCI, og at designprosessen dermed krever nye designtilnærminger (Følstad & Brandtzaeg, 2017). Som en sentral del av omstillingen bør man gå fra å designe et objekt, til å designe en tjeneste. Dette krever en bedre forståelse av brukeren, og man må bevege seg fra å forklare brukeren hva som er tilgjengelig, til å forstå hvordan man kan optimalisere tjenesten til å servere passende løsninger. Samtalen blir imidlertid et objekt som designes, og det blir viktig å tydelig kommunisere bottens muligheter og begrensninger for å gi brukeren realistiske forventninger og forberede brukeren på at botten ikke er feilfri (Yang et al, 2020). Å forberede brukeren på feil bør stå sentralt for å ikke bryte ned brukerens tillit til botten ved avvik (Kocielnik et al. (2019) og Amareshi et al. (2019)). Feil vil nemlig forekomme, både siden teknologien enda ikke kan kommunisere på linje med mennesker, og siden botten er dynamisk og vil gradvis tilpasse seg bruken over tid. Denne uforutsigbarheten er særlig utfordrende å designe for. Tradisjonelt sett har interaksjonsdesignere benyttet seg av visuelle og grafiske representasjoner for å kommunisere med brukere, så overgangen til replikker som chatbottens hovedform for kommunikasjon gjør det nødvendig å tenke annerledes på hvordan man bør kommunisere med brukeren. Brukerne vil imidlertid kommunisere ulikt, og stille like spørsmål ulikt, hvilket skaper utfordringer for designerne.

Hvordan kan G1 og G2 i Amershi et al løse utfordringer i dagens chatbots?
Retningslinje G1 kan skape realistiske forventninger til brukerne, og forberede dem på at chatbotten kan gjøre feil. Dersom dette implementeres riktig minskes sjansen for at brukeren blir frustrert når botten gjør feil, og potensielt migrerer til alternative løsninger i frustrasjon. Eksempler på god implementasjon kan være at botten innledningsvis gir en kort og konsis introduksjon av seg selv og sine egenskaper. Effektiv og tydelig feilhåndtering blir også viktig, slik at brukeren enkelt kan forstå at botten gjør feil, men at den selv er klar over det. Samtidig blir det like viktig at botten gir feedback på at den registrerer og forstår brukerens forespørsler.

Retningslinje G2 bidrar, til likhet med G1, med å gi brukeren realistiske forventninger rundt bottens egenskaper og hvor ofte den gjør feil. Dette er viktig for brukeren tillit til botten, siden det forteller hvor mye man bør stole på svarene man får. Enten bottens feilrate er høy eller lav, er denne åpenheten viktig for å bygge tillit mellom bruker og bot. Åpenheten kan ta form gjennom å bruke historisk data for å vise hvor ofte den svarer riktig på aktuelle spørsmål, og hvor sannsynlig det er at svaret den gir stemmer. Dataen kan genereres ved at brukere selv vurderer bottens svar.

Iterasjon 3
I tilbakemeldingen fra modul 2 fikk jeg skryt for å ha jobbet med tilbakemeldingen fra modul 1. Her savnet retteren min en del kilder, og ønsket å vite hvor jeg hentet informasjonen min fra.
Jeg har derfor supplert med flere kilder i modul 1, og skrevet modul 2 med en mer aktiv kildebruk. Jeg fikk samtidig skryt for å ha valgt et godt AI-infundert system, og for å ha konsekvent referert til systemet gjennom teksten. Med TikTok som mitt valg av AI-infundert system har jeg tilstrebet å aktivt referere til det der det har vært naturlig. Tilbakemeldingen inneholdt imidlertid ingen konstruktiv kritikk. Jeg har likevel gått over teksten, og tatt tak i det jeg selv har funnet av feil og forbedringspotensialer. Dette førte til en generell språkrens, med både mer konsekvent bruk av begreper og mer presis fremlegging av argumenter. Jeg har samtidig forsøkt å fjerne repetisjon de stedene jeg selv har lagt merke til at jeg gjentar meg selv.

Referanser:

Amershi, Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P., 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? ACM International Conference Proceeding Series, 2, 1-5.

Christoforou, E., & Müller, A. (2015). R.U.R. Revisited: Perspectives and Reflections on Modern Robotics. International Journal of Social Robotics, 8(2), 237-246.

Cordeschi. (2007). AI TURNS FIFTY: REVISITING ITS ORIGINS. Applied Artificial Intelligence, 21(4-5), 259–279. https://doi.org/10.1080/08839510701252304

Dautenhahn, K. (2018). Some Brief Thoughts on the Past and Future of Human-Robot Interaction. ACM Transactions on Human-robotic Interaction, 7(1), 1-3.

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

Grudin, J. (2009). AI and HCI: Two fields divided by a common focus. The AI Magazine, 30(4), 48-57.

Haenlein, & Kaplan, A. (2018). A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence. California Management Review, 61(4), 5–14. https://doi.org/10.1177/0008125619864925

Kim, & Lee, K. (2019). A Data-Driven Strategic Model of Common Sense in Machine Ethics of Cares. In Human-Computer Interaction. Perspectives on Design (pp. 319–329). Springer International Publishing. https://doi.org/10.1007/978-3-030-22646-6_23

Kocielnik, Amershi, S., & Bennett, P. (2019). Will You Accept an Imperfect AI? Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 1–14. https://doi.org/10.1145/3290605.3300641

Moran. (2007). Rossum's Universal Robots: Not the Machines. Journal of Endourology, 21(12), 1399–1402. https://doi.org/10.1089/end.2007.0104

Qin, Omar, B., & Musetti, A. (2022). The addiction behavior of short-form video app TikTok: The information quality and system quality perspective. Frontiers in Psychology, 13, 932805–932805. https://doi.org/10.3389/fpsyg.2022.932805

Ruby, Daniel(2022, 19. august). How many TikTok Users Are There?. DemandSage. https://www.demandsage.com/tiktok-user-statistics/

Sabanovic, Milojevic, S., & Kaur, J. (2012). John McCarthy [History]. IEEE Robotics & Automation Magazine, 19(4), 99–106. https://doi.org/10.1109/MRA.2012.2221259

Seising. (2016). Marvin Lee Minsky (1927–2016). Artificial Intelligence in Medicine, 75, 24–31. https://doi.org/10.1016/j.artmed.2016.12.001

Yang, 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