

# 1 Første modul - iterasjon 1

## 1.1 Konsepter, definisjoner, historien til AI og interaksjon med AI

### 1.1.0 Historien om AI

I artikkelen “AI and HCI: Two fields divided by a common focus” skriver Jonathan Grudin om historien til AI helt tilbake til 1950-tallet.

Begrepet *kunstig intelligens* ble først definert av John McCarthy i 1956 og stammer fra matematikk- og ingeniørfaget (Gruding 2009, s 49). Alan Turing skrev følgende i The London Times: *“I do not see why [the computer] should not enter any one of the fields normally covered by the human intellect, and eventually compete on equal terms.”* som illustrerer problemstillingene rundt teknologiens utvikling og bruk av kunstig intelligens (Gruding 2009, s 49).

### 1.1.1 Definisjoner om AI

#### Definisjon 1

Guri Verne og Tone Bratteteig definerer AI i sin forskningsartikkel “Does AI make PD obsolete?” (2018). Dette kan regnes som en nyere definisjon da den første kom allerede i 1956. Verne og Bratteteig tilhører Institutt for Informatikk ved Universitetet i Oslo. Deres definisjon lyder:

*“AI is 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.”* (Bratteteig og Verne, 2018, s.2-3)

Denne definisjonen legger vekt på at kunstig intelligens skal være en form for menneskelig intelligens. Et menneske vil kunne trenge sin intelligens for å utføre den jobben Verne og Bratteteig mener roboten skal klare. Her kommer også et nytt interessant spørsmål inn: Hva gjør dette med skjønn? Mener de at roboten skal ha skjønn når de tar en avgjørelse?

#### Definisjon 2

Wikipedia sin definisjon av kunstig intelligens er interessant å se opp mot andre definisjoner da det er denne siden flertallet av befolkningen bruker om de googler “artificial intelligence”. Denne siden er også høyt i søkemotorikken til google. Deres definisjon lyder:

*“Artificial intelligence (AI) is intelligence demonstrated by machines, as opposed to the natural intelligence displayed by humans or animals.”* (Wikipedia)

I denne definisjonen legges det vekt på at kunstig intelligens er demonstrert av maskiner og har lite med den menneskelige intelligensen å gjøre.

### **Definisjon 3**

John McCarthy definerer AI - i 1956 - som:

*“the science and engineering of making intelligent machines.”* (McCarthy, 1956)

Denne definisjonen legger vekt på hvordan vi kan lage intelligente maskiner. Den går litt i samme baner som Wikipedias definisjon, men har et annet fokus enn Bratteteig og Verne. John McCarthy var en matematiker og logiker, født og oppvokst i USA. Han mottok blant annet Turing-prisen i 1971 for sitt bidrag til kunstig intelligens (Wikipedia).

### **Min definisjon**

Jeg baserer min definisjon på de tre definisjonene over. Det er viktig å ta hensyn til det sosiale samt det tekniske perspektivet på kunstig intelligens. Jeg mener også at definisjonen til Bratteteig og Verne er en fin definisjon da den tar hensyn til det menneskelige i større grad enn de to andre. Min definisjon er derfor:

*“Kunstig intelligens er teknologi som tilegner seg egen kunnskap gjennom læring som kan ligne menneskelig intelligens.”*

### **1.1.2 Artikkel**

Artikkelen “Some Brief Thoughts on the Past and Future of Human-Robot Interaction” av Kerstin Dautenhahn handler om menneske-robot interaksjon (HRI). Dautenhahn skriver om hva hun er spent på, hva hun er bekymret for og hva hun er optimistisk for innenfor HRI (Dautenhahn, 2018).

Hun skriver at HRI handler om langt mer enn bare roboter og at dette er et tverrfaglig felt som handler om for eksempel psykologi og antropologi. Menneske-robot interaksjon utvikles stadig, men for at det skal utvikles i riktig retning må man også ta i bruk den riktige metodikken og de riktige feltene for å svare på forskningsspørsmålene. Dautenhahn er av denne grunn også bekymret for hvordan HRI kommer til å utvikles over tid. Hun nevner blant

annet definisjonen av hva en robot er, men hva er egentlig en robot? Hun nevner også at menneske-robot interaksjon er veldig forskjellig fra menneske-menneske interaksjon, som gjør hele forskningen utfordrende. Det brukes ofte eksperimentelle tilnærminger til HRI i arbeidet med roboter. Slikt arbeid utelater hvordan vi egentlig interagerer med en robot når vi står ansikt-til-ansikt med den i det virkelige liv.

Artikkelen tar for seg fine poenger og drøftinger, men jeg lurer fortsatt på følgende:

- Er det noen hendelser/eksempler på hva som har gjort Dautenhahn skeptisk til HRI?
- På hvilken måte kan ulike forskningsfelt brukes i studiet av HRI?

### **1.1.3 Et moderne selskap og bruk av AI**

Et spesielt selskap som fanget min interesse i sin fremstilling av AI var Posten. Et eksempel på bruk av AI hos Posten er leserrateforbedring i forbindelse med maskinell brevsortering, samt chatbots (Posten, 2020) . De henviser til Regjeringens utsagn om at for få mennesker kan AI, som gir de få personene mye makt i å påvirke vår hverdag i fremtiden da AI utvikler seg hele tiden. Posten oppfordrer “mannen i gata” til å ta kurs og lære mer om AI.

### **1.1.4 Black Mirror**

Black Mirror er en serie som er kjent for å se de mørke sidene av teknologi. Jeg skal skrive litt om episode 1 sesong 2. En kvinnes kjæreste dør i en tragisk bilulykke mens hun er gravid. Hun kjøper derfor en robot, basert på AI, som erstatter kjæresten sin. Roboten har det samme utseende og stemme, men mangler de personlighetstrekkene og små tingene som gjorde kjæresten til den han var. Kvinnen føler seg derfor ikke noe bedre av denne roboten i det lange løp og bestemmer seg for å kvitte seg med den - hun legger den på loftet.

## **1.2 Roboter og AI systemer**

### **1.2.0 Ordet robot**

Karel Čapek, en tsjekkisk skribent, introduserte ordet Robot i sitt teaterstykke “Rossum’s Universal Robots” i 1920. Ordet stammer fra et kristent slovensk miljø og ordet *Robata* som betyr “forced labor” (Science Friday, 2011).

### **1.2.1 Definisjoner**

#### **Definisjon 1**

I 1979 definerte Robot Institute of America roboten som

*“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”* (Thrun, 2004).

Denne definisjonen tar utgangspunkt i at en robot er en “ting” som kan fysisk flytte på gjenstander og utføre en rekke oppgaver. De fokuserer i hovedsak på robotens bruksegenskaper.

## **Definisjon 2**

Merriam Webster’s leksikon definerer roboten i 1993:

*“An automatic device that performs functions normally ascribed to humans or a machine in the form of a human.”* (Thrun, 2004).

Denne definisjonen plasserer egenskapene til roboten ved siden av menneske. Den tar i større grad utgangspunkt i at roboten skal utføre menneskelige oppgaver enn hva den første definisjonen gjorde.

## **Min definisjon**

Basert på definisjonene over er min definisjon følgende:

*“En robot er en automatisk maskin og et objekt som er i stand til å utføre fysiske oppgaver.”*

Denne definisjonen tar utgangspunkt i at roboten kan utføre både menneskelige oppgaver i form av en chatbot som svarer på dine spørsmål, samt fysiske oppgaver som å flytte gjenstander.

### **1.2.2 AI og Robot**

Med utgangspunkt i definisjonene om AI og robot tenker jeg at den største forskjellen mellom dem er “forståelse” og “utseende/form”. En robot, basert på definisjonene, er et fysisk objekt som skal gjennomføre programmerte oppgaver. AI derimot trenger ikke være noe fysisk du kan ta på eller utføre fysiske oppgaver. Den tilegner seg kunnskap gjennom oppgaver.

Likhetene mellom de er at roboten kan utføre menneskelige oppgaver, hvor AI også lærer seg hvordan mennesket oppfører seg.

### **1.2.3 Robotgressklipper**

En vanlig robot er “robotgressklipperen”. Denne roboten følger en kabel som legges ned i plenen som gjør at roboten ikke går utenfor dette området. Man interagerer med den gjennom en app som instruerer blant annet hvor lenge og hvor ofte roboten skal klippe.

## **1.3 Universell design og AI systemer**

### **1.3.1 Definisjon**

National Disability Authority definerer universal design som:

*“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.”* (National Disability Authority).

Definisjonen mener at design av et “miljø” må være tilpasset alle brukergrupper, uavhengig av hvilke funksjonshemninger man har. Derfor mener jeg at det må legges tilstrekkelig grunnlag for universal design i en designprosess. På denne måten inkluderer man alle fra start og slipper uforutsette hendelser underveis. Det er også viktig å kartlegge hvem som skal bruke det man designer.

### **1.3.2 AIs potensiale**

AI kan bli viktig for å legge tilrettelegge for de med funksjonshemninger. I dag kan man for eksempel bruke en robotgressklipper og styre den fra telefonen. Man kan også ha en robotstøvsuger på lik linje som gressklipperen. Disse AI robotene gjør at man kan få utført oppgaver man kanskje ellers ikke ville klart.

### **1.3.3 Inkludering og ekskludering**

Ansiktsgjenkjenning har vist seg å være ekskluderende da den ikke detekterer alle mennesker, avhengig av etnisk bakgrunn. Dette har skjedd fordi AIen er trent opp til å gjenkjenne ansikter basert på bilder av millioner av mennesker. Det er imidlertid tatt for lite hensyn til ulike etnisiteter (Time, 2019).

Et eksempel på inkludering er roboter. Gressklipper, selvkjørende biler og støvsuger kan gi muligheter til de som ikke klarer å utføre slike oppgaver i det daglige.

### **1.3.4 Forstå og Forståelse - skjønner maskiner?**

Jeg tolker begrepet “forstå” som at AI har en kunnskapsgrunnlag for hva den skal gjøre i en gitt situasjon. Begrepet “forståelse” er mer kompleks da jeg tolker dette som en slags menneskelig intelligens som handler om skjønn. Maskiner kan forstå, men har nok en lavere grad av forståelse. Et eksempel kan være psykologi eller saksbehandlere. Ved å sette en robot til disse oppgavene mister man den grunnleggende menneskelige medfølelsen og forståelsen for situasjonen.

## **1.4 Guideline for Human-AI interaction**

### **1.4.1 AI guidelines**

Jeg har valgt retningslinje nummer 11 “*Make clear why the system did what it did.*”. Denne retningslinjen forteller at brukeren trenger en forklaring på hvorfor systemet og AI oppfører seg som den gjør. Et eksempel på dette er ved å gi brukeren feedback underveis - enten når noe går galt og/eller underveis for å få bekreftelse på at man gjør ting riktig.

### **1.4.2 AI og HCI**

HCI handler i hovedsak om brukeren, mens AI handler mer om systemenes oppførsel. Ved å ta utgangspunkt i Jakob Niensens 10 heuristikker ser man at det finnes mange likheter til AI-prinsippene (Nielsen, 1994). Eksempler er for eksempel at begge fokuserer på brukervennlighet og kontroll over systemene.

## 2 Andre modul - iterasjon 2

### 2.1 Karakteristikker ved “AI-infused” systemer

#### 2.1.1 Systemer med AI

I den første forelesningen av modul 2 med Asbjørn Følstad ble det presentert fire karakteristikk ved AI-baserte systemer: Læring, forbedring, black box og at systemet drives av store datasett.

Av plasshensyn beskriver jeg kun noen av karakteristikkene fra artiklene i sammenheng med karakteristikkene fra forelesning.

##### *2.1.1.1 Læring*

Den første karakteristikken refererer til et dynamisk system som endrer seg basert på hvem som bruker systemet. Ved at AI-baserte systemer lærer basert på brukere så vil systemet også forbedre resultatene sine og ha mindre feilmargin. Amershi et al. skriver at siden AI-baserte systemer lærer over tid kan dette føre til inkonsekvente komponenter (Amershi et al., 2019).

##### *2.1.1.2 Forbedring*

Knyttet til delkapittelet om læring så skal AI også forbedre seg over tid. Amershi et al. skriver at feil (“error”) er vanlig i AI-baserte systemer og at at forebygging av feil er vanskelig å oppnå på en pålitelig måte. Dette har også bidratt til økende arbeid med forklaringer og tolkninger av AI. På denne måten vil man støtte menneskelig verifisering og forhåpentligvis redusere sannsynligheten for potensielt farlige og kostbare utfall (Amershi et al., 2019).

##### *2.1.1.3 Black box*

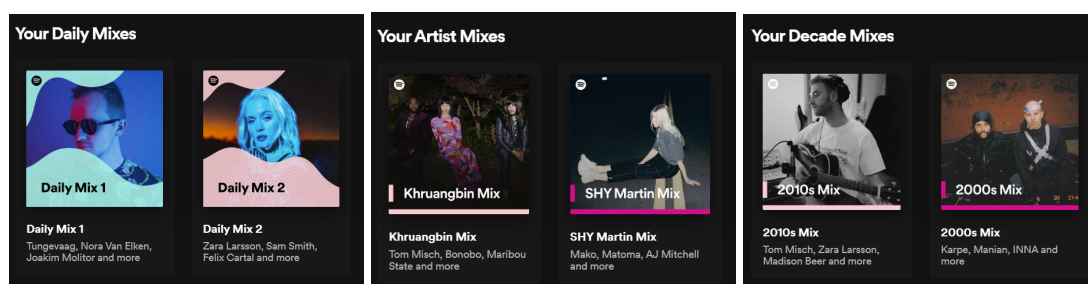
Black box er et AI system hvor input og funksjonaliteten til systemet ikke kan ses av brukere (Følstad, 2021, forelesning). Kocielnik et al. skriver at AI-baserte systemer ofte ikke leverer perfekt data, mens brukerne forventer det motsatte. Dette kan også påvirkes av systemets bruk. I artikkelen skriver de også at man kan gjøre brukerne klar over dette på forhånd før kjøp, men at dette vil gi brukerne negative tanker om systemet (Kocielnik et al., 2019).

### 1.1.4 Store datasett

AI-baserte systemer drives av store datasett gjennom interaksjon med brukerne (Følstad, 2021, forelesning). Yang et al skriver at den negative siden med datasett er at den nye dataen kan føre systemet i feil retning. Den ideelle situasjonen er at systemet skal vokse og integreres med den nye dataen fra brukere, noe som da ikke alltid skjer (Yang et al., 2020, s. 6-7).

## 1.2 Spotify og AI

Et system som jeg og millioner av andre bruker, er Spotify. Denne musikkapplikasjonen bruker AI blant annet til å gi deg personlige forslag til sanger, slik som Daily Mix, Your Artist Mixes and Your Decade Mixes. Spotify gjenkjenner hvilke sanger du hører mest på og kan gi deg forslag til lignende sanger, samt at den finner ut hvilke artister du hører mest på og hvilke sanger du har hørt mest på fra ulike tiår. Dette kan karakteriseres som “læring”.



Som bruker av Spotify påvirker karakteristikken “læring” oss veldig. Man får et mer personlig forhold til musikkapplikasjonen i tillegg til at du lettere kan finne nye sanger som du liker.

## 2.2 Menneske-AI interaksjonsdesign

### 2.1.1 Sammendrag

Amershi et al., skriver “*AI-infused systems can violate established usability guidelines of traditional user interface design.*” (Amershi et al., 2019). AI-baserte systemer lærer over tid og kan derfor være inkonsekvente og reagere forskjellig avhengig av ytre faktorer som støy og lys. Forfatterne skriver videre at designere og utviklere fortsatt sliter med å lage intuitive og effektive AI-baserte systemer, spesielt siden denne teknologien stadig er i utvikling. De



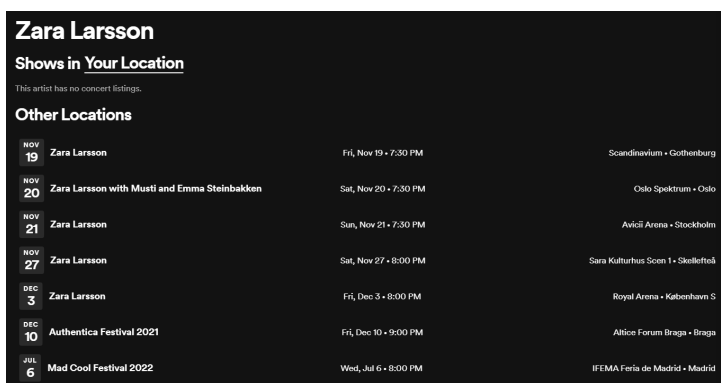
foreslår derfor 18 delte retningslinjer som kan gjenbrukes og hjelpe med design og evaluering av AI systemer (Amershi et al., 2019).

I motsetning til Amershi et al., som ønsker å forbedre designet i AI-baserte systemer, så ønsker Kocielnik å styre forventningene til sluttbrukeren (Kocielnik et al., 2019). De foreslår tre teknikker for å sette forventninger: 1) Nøyaktighetsindikator, 2) Forklaring basert på eksempler, og 3) Ytelseskontroll. De tre teknikkene skal brukes til å bevare brukertilfredsheten og at det aksepteres av sluttbrukerne at systemet ikke er helt fullkomment (Kocielnik et al., 2019).

### 2.1.2 Retningslinjer

#### Retningslinje 4 - Show contextually relevant information.

Hvis jeg søker opp for eksempel Zara Larsson på Spotify så får jeg opp når hun har konsert nærme meg.

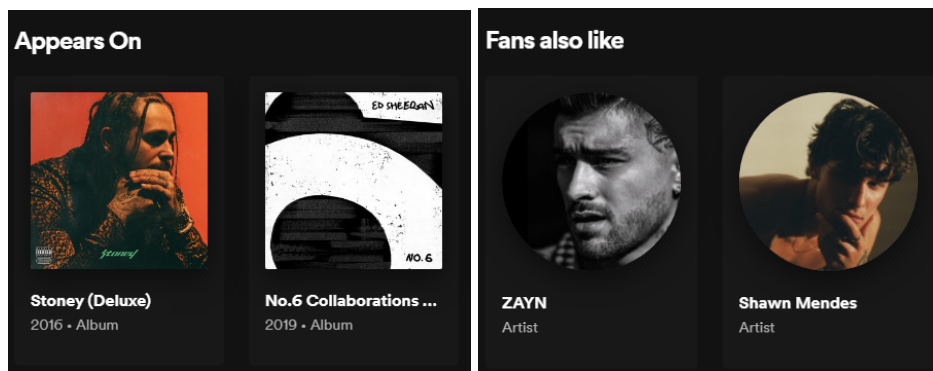


The screenshot shows the Spotify artist page for Zara Larsson. It features a section titled "Shows in Your Location" which states "This artist has no concert listings." Below this, there is a section titled "Other Locations" with a list of upcoming concerts. Each entry includes the date, the artist's name, the time, and the venue.

Date	Artist	Time	Venue
NOV 19	Zara Larsson	Fri, Nov 19 • 7:30 PM	Scandinavium • Gothenburg
NOV 20	Zara Larsson with Musti and Emma Steinbakken	Sat, Nov 20 • 7:30 PM	Oslo Spektrum • Oslo
NOV 21	Zara Larsson	Sun, Nov 21 • 7:30 PM	Avicii Arena • Stockholm
NOV 27	Zara Larsson	Sat, Nov 27 • 8:00 PM	Sara Kulturhus Scen 1 • Skellefteå
DEC 3	Zara Larsson	Fri, Dec 3 • 8:00 PM	Royal Arena • København S
DEC 10	Authentica Festival 2021	Fri, Dec 10 • 9:00 PM	Altice Forum Braga • Braga
JUL 6	Mad Cool Festival 2022	Wed, Jul 6 • 8:00 PM	IFEMA Feria de Madrid • Madrid

#### Retningslinje 12 - Remember recent interactions.

Hvis man søker på en artist på Spotify, så vil du få opp en liste hvor det står “medvirkende”. Spotify har funnet alle album som denne artisten tilhører. I tillegg vil man få opp “lignende” artister. Søker du opp for eksempel, Justin Bieber, så får du forslag til artister i samme kategori.



### 2.1.3 Sammendrag

Bender et al. skriver i sin artikkel hvilke kostnader og risikoer som kommer med jaget etter større "language models" (LM) som blant annet pengekonstanter og miljøkostnader.

Energien det kreves for å holde i gang disse datasentrene er så store at miljøutslippene kan sammenlignes med et amerikansk charterfly (Bender et al., 2021, s. 612).

Forfatterne oppfordrer forskerne å nøye vurdere om fordelene veier opp for risikoene. De oppfordrer også fagfeltet til å innse at applikasjoner som tar sikte på å etterligne mennesker kan medføre risiko for ekstreme skader (Bender et al., 2021, s. 619).

## 2.3 Chatbots/samtale brukergrensesnitt

### 2.3.1 Hovedutfordringer ved å designe en chatbot

Jeg tenker at det finnes en rekke etiske utfordringer rundt design av en chatbot, samt kompleksiteten og hvordan mennesker vanligvis snakker med hverandre. Et eksempel kan være når man snakker med for eksempel en chatbot som er kundebehandler. Personlig ønsker jeg ofte å få snakke med mennesket så fort som mulig og ikke chatboten, fordi mennesket ofte gir meg mer troverdige svar.

Følstad og Brandtzæg adresserer tre ulike utfordringer når det kommer til design av chatbots. Den første handler om at chatbots regner med at vi som brukere skriver inn riktige ord og uttrykk. Forfatterne legger derfor vekt på at vi må legge designfokus på samtalen og ikke på det grafiske uttrykket (Følstad og Brandtzæg, 2017).

Den andre utfordringen mener Følstad og Brandtzæg er at vi må legge fokuset over på service design. På sikt brukes kanskje chatbots til flere tjenester enn kun en meldingstjeneste. De oppfordrer derfor til å se på hele tjenesten når man designer en chatbot (Følstad og Brandtzæg, 2017).

Den siste utfordringen deres er chatbotens evne til å interagere med et nettverk av mennesker og smarte maskiner. Disse ulike nettverkene av interaksjon må man tenke på når man designer en chatbot (Følstad og Brandtzæg, 2017).

### 2.3.2 Hovedutfordringer ved å designe en chatbot

	AI Design Guidelines		Example Applications of Guidelines
Initially	G1	<b>Make clear what the system can do.</b> Help the user understand what the AI system is capable of doing.	[Activity Trackers, Product #1] "Displays all the metrics that it tracks and explains how. Metrics include movement metrics such as steps, distance traveled, length of time exercised, and all-day calorie burn, for a day."
	G2	<b>Make clear how well the system can do what it can do.</b> Help the user understand how often the AI system may make mistakes.	[Music Recommenders, Product #1] "A little bit of hedging language: 'we think you'll like.'"

Retningslinje 1: I delkapittel 2.2.1 skrev jeg om at Kocielnik et al. ønsker å styre forventningene til brukeren (Kocielnik et al., 2019). Jeg tenker at dette også er viktig for chatbots slik at man som bruker hele tiden har en forståelse av "hvem man prater med".

Retningslinje 2: Denne retningslinjen er litt tilsvarende den første, men legger mer vekt på hvor bra et system utfører jobben sin. Et eksempel er ulike bankers chatbot. De kan ikke svare på de større spørsmålene, men kan gi deg svar på konkrete ting. Dette handler også om å forventningsstyre sluttbruker.

## 3 Appendix - tilbakemeldinger

Basert på tilbakemeldingene jeg fikk på iterasjon 1 har jeg endret forklaringen på min definisjon av "robot" i delkapittel 1.2.1. Jeg har gjort den noe lenger og litt tydeligere.

## Litteraturliste

Amershi, S. Weld, D. Vorvoreanu, M. Fourney, A. Nushi, B. Collisson, P. Teevan, J. (2019). *Guidelines for human-AI interaction*. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. ACM. Lest: 15.10.2021

Bender, E. M. Gebru, T. McMillan-Major, A. Mitchell, M. (2021). *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?*. In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. ACM. Lest: 19.10.2021

Buolamwini, Joy. (07.02.2019) *Artificial Intelligence Has a Problem With Gender and Racial Bias. Here's How to Solve It*. Time Magazine. Hentet fra: <https://time.com/5520558/artificial-intelligence-racial-gender-bias/> Lest: 10.09.21

Bratteteig, Tone. Verne, Guri. (2018). *Does AI make PD obsolete? Exploring challenges from Artificial Intelligence to Participatory Design*.

Dautenhahn, K., (2018). *Some Brief Thoughts on the Past and Future of Human-Robot Interaction*.

Følstad, A. Brandtzæg, P. B. (2017). *Chatbots and the new world of HCI. interactions*, 24(4), 38-42. Lest: 19.10.2021

Grudin, Jonathan. (2009). *AI and HCI: Two Fields Divided by a Common Focus Copyright*. Association for the Advancement of Artificial Intelligence.

Kocielnik, R. Amershi, S. Bennett, P.N. (2019). *Will You Accept an Imperfect AI?: Exploring Designs for Adjusting End-user Expectations of AI Systems*. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems ACM. Lest: 15.10.2021

National Disability Authority. *What is Universal Design?*  
Hentet fra: <http://universaldesign.ie/What-is-Universal-Design/> Lest: 10.09.21

Nielsen, Jakob. (1994). *10 Usability Heuristics for User Interface Design*. Oppdatert 15.11.2020.

Hentet fra: <https://www.nngroup.com/articles/ten-usability-heuristics/> Lest: 10.09.2021

Posten. (07.05.2020). *Posten støtter KI-løftet*.

Hentet fra: <https://www.postennorge.no/presse/arkiv/posten-stotter-ki-loftet> Lest: 09.09.21

Thrun, Sebastian . (2004). *Toward a Framework for Human–Robot Interaction*. Stanford University.

Science Daily. *Artificial Intelligence*. Hentet fra:

[https://www.sciencedaily.com/terms/artificial\\_intelligence.htm](https://www.sciencedaily.com/terms/artificial_intelligence.htm) Lest: 08.09.2021

Wikipedia. *Artificial intelligence*.

Hentet fra: [https://en.wikipedia.org/wiki/Artificial\\_intelligence](https://en.wikipedia.org/wiki/Artificial_intelligence) Lest: 08.09.2021

Wikipedia. *John McCarthy*.

Hentet fra: [https://no.wikipedia.org/wiki/John\\_McCarthy](https://no.wikipedia.org/wiki/John_McCarthy) Lest: 08.09.2021

Yang, Q. Steinfeld, A. Rosé, C. Zimmerman, J. (2020). *Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design*. In Proceedings of the 2020 CHI conference on human factors in computing systems. Lest: 15.10.2021