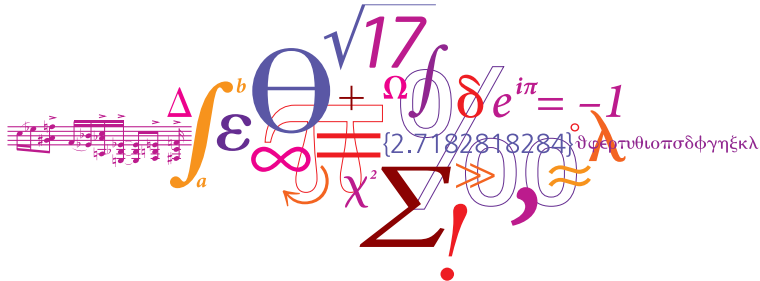


# Muligheder og udfordringer i kunstig intelligens

Thomas Bolander, DTU Compute

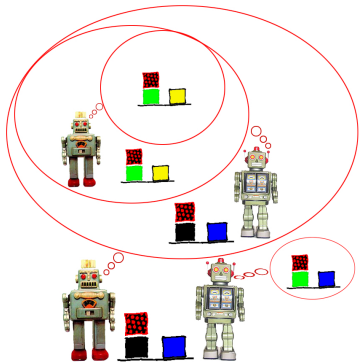
ATV, 4. maj 2018



# Lidt om mig selv

## Thomas Bolander

- Lektor i **logik** og **kunstig intelligens** ved **DTU Compute, Danmarks Tekniske Universitet**.
- Medlem af **SIRI-kommission**, nedsat af Ida Auken og Ingeniørforeningen i Danmark (IDA).
- **Aktuel forskning**: Sociale aspekter af kunstig intelligens. At udstyre kunstig intelligens-systemer med en **Theory of Mind**.
- Medarrangør og videnskabelig rådgiver for **Science & Cocktails** ([scienceandcocktails.org](http://scienceandcocktails.org)).



# Hvad er kunstig intelligens?

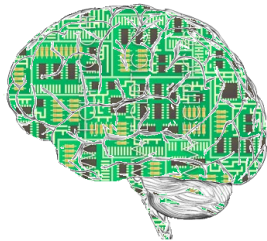
John McCarthy, den kunstige intelligens' fader, definerer:

*“Artificial intelligence is the **science and engineering of making intelligent machines, especially intelligent computer programs.**”*

**Problem:** Utroligt mange forskellige former for intelligens, og på meget forskellige niveauer.

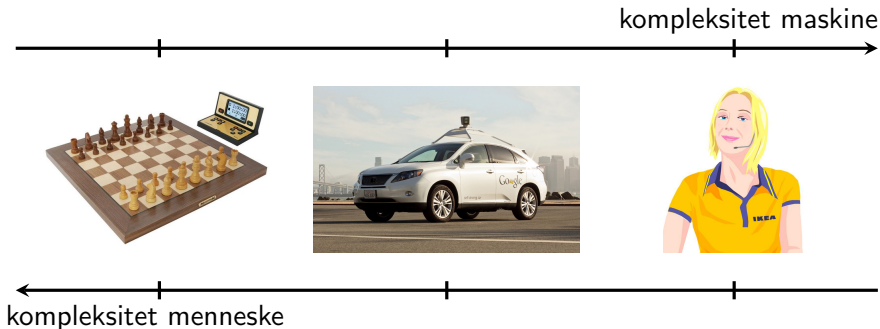


John McCarthy, 2006



# Karakteristika ved kunstig intelligens-systemer i dag

- **Specialiserede systemer:** Løser veldefinerede, velafgrænsede problemer.
- **Revolutionen skyldes i høj grad beregningskraft og data:** mere end udvikling af nye metoder med højere kognitive evner.





# To hovedparadigmer i kunstig intelligens

**Det symbolske paradigme (1950–):** Simulerer menneskelig symbolsk, sproglig, bevidst ræsonnering. Søgning, planlægning, logisk ræsonnering. **Eks:** skakcomputer.



robust, forudsigelig, forklarlig



nøje afgrænsede evner



fleksibel, læring



aldrig 100% forudsigelig/fejlfri

**Det subsymbolske paradigme (1980–):** Simulerer de fundamentale fysiske processer i den menneskelige hjerne. Neurale netværk. **Eks:** billedgenkendelse.

symbolsk



subsymbolsk

## Lidt subsymbolisk magi...



(Liu et al.: Unsupervised Image-to-Image Translation Networks, NIPS 2017)

## Asociale hospitalsrobotter

Frustrerede brugere af hospitalsrobotter i USA:

- *"I'm on the phone! If you say 'TUG has arrived' one more time I'm going to kick you in your camera."*
- *"It doesn't have the manners we teach our children. I find it insulting that I stand out of the way for patients... but it just barrels right on."*



*TUG hospitalsrobot*

(Colin Barras, New Scientist, vol. 2738, 2009)

## Test af Aida (Amelia), SEB

Dialog mellem mig og Aida på SEBs hjemmeside, 29. maj 2017:

**Thomas:** Hur öppnar jeg ett konto?

**Aida (Amelia):** Är du kund hos oss sedan tidigare?

**Thomas:** Nej.

**Aida:** För at skaffa Privatkonto behöver du vara kund i SEB. Vill du at jeg guidar dig?

**Thomas:** Ja tak.

**Aida:** För att kunne hitta den enklaste løsningen for dig behöver jeg stille några frågor. Är du over 18 år?

**Thomas:** Ja.

**Aida:** Kan du legitimere dig med mobilt BankID?

**Thomas:** Jag är från Danmark.

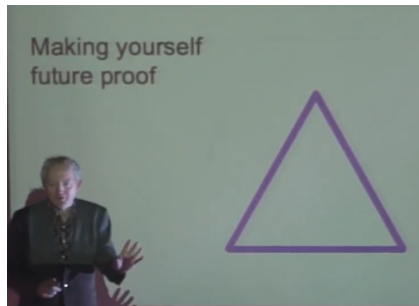
**Aida:** Jag förstår inte riktigt vad du vill ha hjælp med, men formulera dig gärna på ett annat sätt.



## De 3 sværeste problemer i kunstig intelligens



Carl Frey, 20. april 2017  
Koldingfjord, Kolding



Toby Walsh, 18. marts 2017  
Science & Cocktails, København

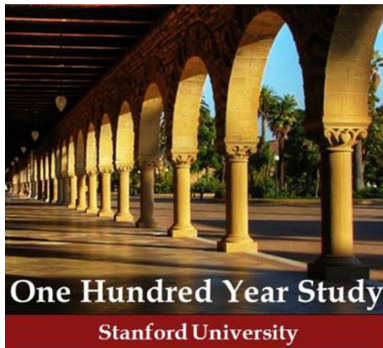
Begge har **social intelligens** blandt de 3 menneskelige kognitive evner som er sværest at simulere i en computer eller robot.

**Social intelligens:** Evnen til at forstå andre og den sociale kontekst effektivt og dermed være i stand til at interagere succesfuldt med andre.

# Kunstig intelligens, mennesker og tillid

*AI technologies already pervade our lives. As they become a central force in society, the field is shifting from simply building systems that are intelligent to building intelligent systems that are **human-aware and trustworthy**.*

(Min fremhævning)



(One Hundred Year Study on AI: 2015–2016, Stanford University, 6. september 2016)

# Tillid og forklarlighed

Hvornår har vi tillid til de **beslutninger**, **forudsigelser** eller **klassifikationer** et kunstig intelligens-system foretager:

1. Når systemet **aldrig tager fejl?**
2. Når systemet **meget sjældent tager fejl?**
3. Når systemet **for det meste ikke tager fejl**, men når det gør det, så kan det give en acceptabel og **forklarlig årsag** til at tage fejl.



## Forklarlig kunstig intelligens (explainable AI)

- Uheld med førerløse biler og racistisk kunstig intelligens kræver mere **transparent** og **forklarlig** kunstig intelligens.
- Neurale netværk og maskinlæring (**subsymbolsk**) er naturligt **intransparente** (black box).
- Regelbaserede systemer (**symbolsk**) er naturligt **transparente**, men er ofte urealistiske at opbygge manuelt.
- Bedste nuværende bud er at **kombinere**: outputtet af maskinlæring (subsymbolsk) er regler og eksplicite modeller (symbolsk) som kan inspiceres, forklares til, forstås af og modificeres af mennesker. I tråd med DARPA's "Third wave of AI" og meget nuværende forskning hos bl.a. Google DeepMind.

Big Data-mantraet omkring "hvad, ikke hvorfor" (f.eks. Viktor Mayer-Schonberger and Kenneth Cukier: Big Data—A Revolution that will Transform how we Live, Work and Think) er udfordret når beslutninger tages af Big Data-algoritmer, og de påvirkede mennesker har **ret til forklaring** (jvf. GDPR).