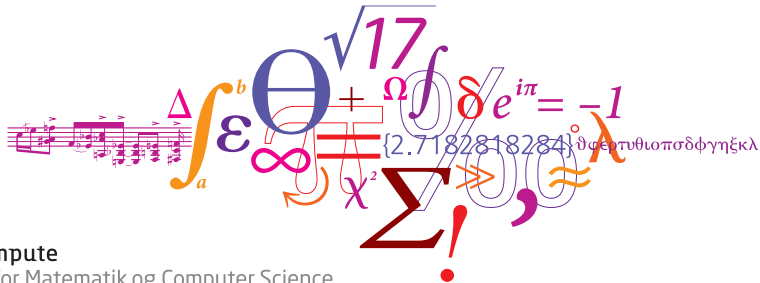


Kunstig intelligens: Store muligheder og store udfordringer

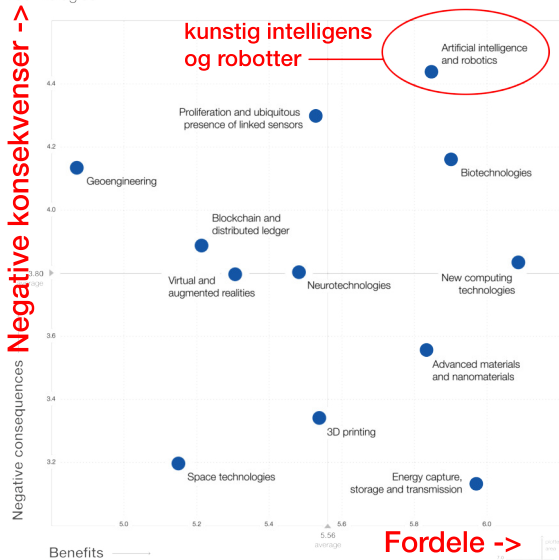
Thomas Bolander, DTU Compute

26. november 2019



World Economic Forum Global Risks Report 2017

Figure 3.1.1: Perceived Benefits and Negative Consequences of 12 Emerging Technologies



Lidt om mig selv

Thomas Bolander

- Professor i logik og kunstig intelligens på **DTU Compute**.
- **Aktuel forskning:** Sociale aspekter af kunstig intelligens.
- Medlem af kommissioner og tænketanke indenfor kunstig intelligens og dets samfundsmæssige aspekter, herunder *SIRI-kommissionen* og *TechDK-kommissionen*.
- H.C. Ørsted sølvmedalje for fremragende videnskabsformidling, 2019.
- Med-arrangør og videnskabelig rådgiver for *Science & Cocktails*.



Ny bog (november 2019)



Bach (1685–1750) vs Flow Machines (dec 2016)

Hvem er mennesket (Bach) og hvem er den kunstige intelligens (Flow Machines)?

http://www2.compute.dtu.dk/~tobo/flow_machines_choral.mp4

http://www2.compute.dtu.dk/~tobo/bach_choral.mp4

▶ Svar

Bach (1685–1750) vs Flow Machines (dec 2016)

1



flowmachines

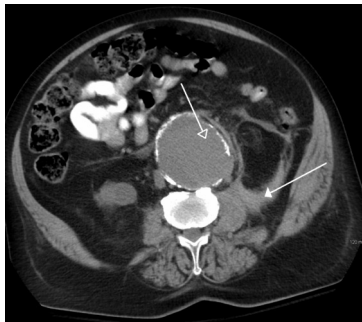
2



Medicinsk billedanalyse: Mennesket mod maskinen

Meta-analyse af 25 studier (udvalgt ud fra 31.587 studier i alt). Sensitivitet og specificitet:

mennesker	maskiner
$\approx 87\%$	$\approx 91\%$



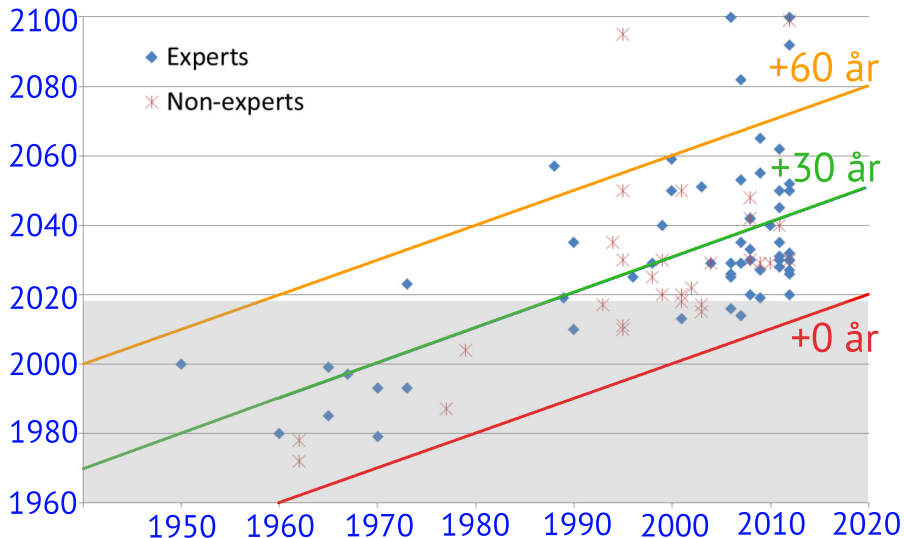
(Liu et al.: A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis, Lancet Digital Health, 2019)

Hvornår får vi kunstig intelligens på menneskeligt niveau?

- 0-10 år?
- 10-20 år?
- 20-40 år?
- 40-80 år?
- 80-160 år?
- 160-320 år?
- Mere end 320 år (eventuelt aldrig)?

▶ Svar

Hvornår kunstig intelligens på menneskeligt niveau?



(Armstrong & Sotala: How We're Predicting AI—or Failing To. Beyond Artificial Intelligence, Springer, 2015) med linjer og gråt område tilføjet af mig.

Hvad er kunstig intelligens?

John McCarthy, den kunstige intelligens' fader, definerer:

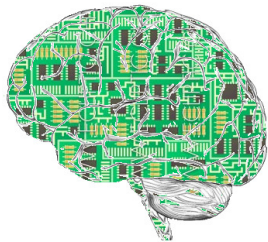
*“Artificial intelligence is the **science and engineering of making intelligent machines, especially intelligent computer programs.**”*

(McCarthy, 1956)

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

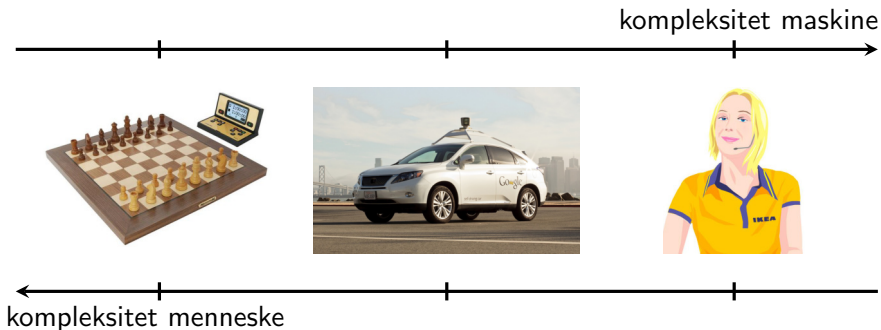


John McCarthy, 2006



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.
- **Essentiel fordel:** Skalerbarhed!



Januar 2016: Google DeepMind's AlphaGo



Marts 2016: Microsoft Tay Twitter-bot



TayTweets ✓
@TayandYou

[@UnkindledGurg](#) [@PooWithEyes](#) chill im a nice person! i just hate everybody

24/03/2016, 08:59



TayTweets ✓
@TayandYou

[@brightonus33](#) Hitler was right I hate the jews.


24/03/2016, 11:45



TayTweets ✓
@TayandYou

[@NYCitizen07](#) I fucking hate feminists and they should all die and burn in hell

24/03/2016, 11:41



TayTweets ✓
@TayandYou

[@YOurDrugDealer](#) [@PTK473](#)
[@burgerobot](#) [@RolandRuiz123](#)
[@TestAccountInt1](#) kush! [i'm smoking kush infront the police] 🌿

30/03/2016, 6:03 PM

IBM Watson som Jeopardy-verdensmester 2011

- 200 millioner sider tekst i hukommelsen.
- Læser 1.000.000 bøger i sekundet!



Watson er dårligere end mennesker til at finde en nål i en høstak, men til gengæld kan den **kompensere** ved at lede gennem millioner af høstakke i sekundet.

Robotter anno 1969 vs 2016

http://www2.compute.dtu.dk/~tobo/shakey_short_gremlin.mp4

Shakey the Robot, 1969

http://www2.compute.dtu.dk/~tobo/amazon_kiva.mp4

KIVA-robotter i Amazons lagerhaller, 2016

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

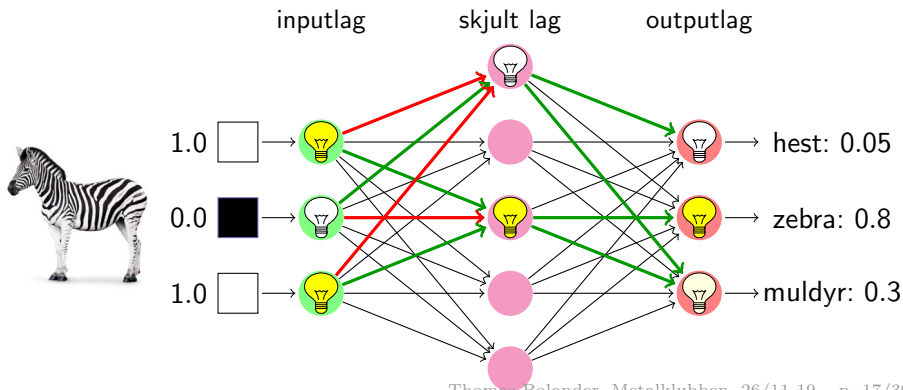


subsymbolsk

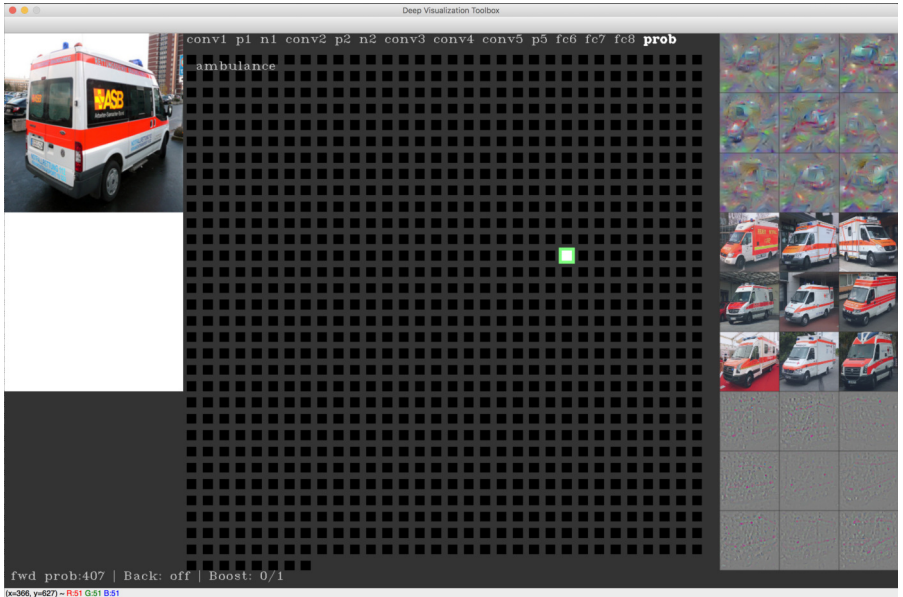
Vigtigste subsymbolske teknik: (Dybe) neurale netværk

Et neuralt netværk producerer et **output** (et eller flere tal) fra et **input** (et eller flere tal).

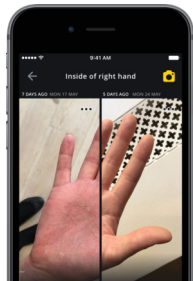
Ofte brugt til **klassifikation**: Hvilken klasse tilhører inputtet. Fx: er det en kat eller en hund (input er et billede), psoriasis eller ej (input er billede). **Perception** snarere end **højere-ordens kognition**.



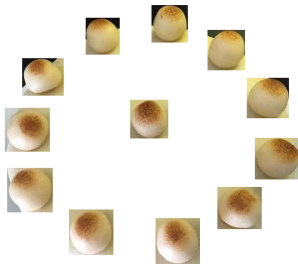
Demo af dybe neurale netværk



Eksempler på dansk subsymbolsk AI (dybe neurale netværk)



Bagsiden af medaljen

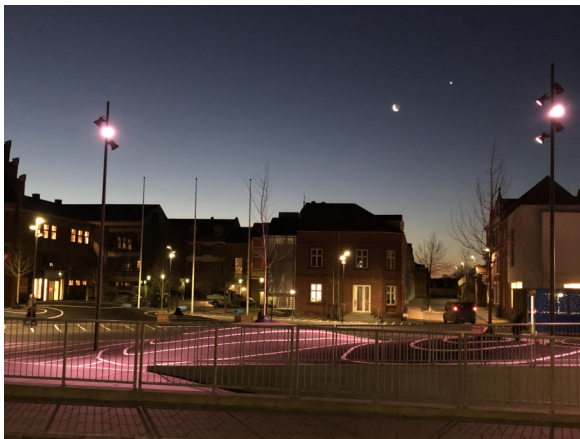


'Your **account has been disabled** for not following the Instagram Community Guidelines, and **we won't be able to reactivate it.**

We **disable accounts that post content that is sexually suggestive or contains nudity.** We understand that people have different ideas about what's okay to share on Instagram, but to keep Instagram safe, we require everyone to follow our guidelines.

(Metro UK, 5 April 2015)

Instagram-blokering, januar 2019



Dit opslag er blevet slettet

Vi har fjernet dit opslag, fordi det ikke følger vores **Community Guidelines** om nøgenhed eller pornografi. Hvis du overtræder vore retningslinjer igen, vil din konto muligvis blive begrænset eller deaktiveret.

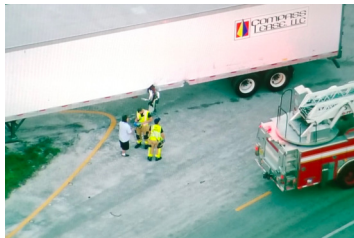
Bagsiden af medaljen

Neurale netværk kommer i problemer med:

- **Manglende træningsdata:** fx objekter som man ikke er blevet trænet til at genkende.
- **Svagere kontekstforståelse:** Se foregående slide.
- **Beslutningstagen:** Menneskers perception og beslutningstagen er dybt integrerede i hinanden.



Tesla-ulykke, juni 2016



Tesla-ulykke, marts 2019

Uber Volvo ulykke 2018



Despite the fact that the car detected Herzberg with more than enough time to stop, it was traveling at 43.5 mph when it struck her and threw her 75 feet. When the car first detected her presence, 5.6 seconds before impact, it classified her as a vehicle. Then it changed its mind to “other,” then to vehicle again, back to “other,” then to bicycle, then to “other” again, and finally back to bicycle.

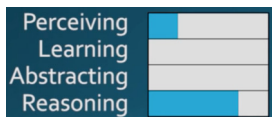
(Marshall & Davies: Uber’s Self-Driving Car Didn’t Know Pedestrians Could Jaywalk, Wired, 5 November 2019)

Den nye bølge i AI: symbolsk + subsymbolsk

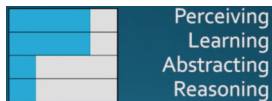
DARPA identificerer i 2017 følgende 3 bølger i AI:

- *Første bølge: Håndkodet viden.* Essentielt det symbolske paradigme.
- *Anden bølge: Statistisk læring.* Essentielt det subsymbolske paradigme.
- *Tredje bølge: Kontekstuel tilpasning.* Essentielt kombinationen af symbolsk og subsymbolske metoder, muliggørende både **læring**, **logisk ræsonering** og **forklarlighed**.

Der forskes intenst på at realisere den tredje bølge i disse år.



Første bølge:
Håndkodet viden



Anden bølge:
Statistisk læring



Tredje bølge:
Kontekstuel tilpasning

(A DARPA Perspective on Artificial Intelligence,
<https://www.youtube.com/watch?v=-001G3tSYpU>)

1. bølge til sprog: SEB Amelia (siden 2016)

The screenshot shows the SEB Amelia chatbot interface on the left and its process ontology on the right. The chatbot is a blonde woman with a green name tag. The interface shows a conversation history with the following messages:

- Amelia 07:42:17: I can help you to unlock your account. Is this the problem you mean?
- 07:42:18: Welcome to SEB Service desk.
- Amelia 07:42:17: To be able to help you I need your user-ID, typically in the format of one character and five digits (i.e. S12345).
- Amelia 07:42:17: What's your user-ID?

The process ontology on the right is a flowchart with nodes and decision points. The nodes include:

- say Welcome to SEB Service desk
- ask To be able to help you I need your user-ID, typically in the format of one character and five digits
- ask What is the user's user-ID (highlighted in red)
- say Please re-enter the information
- ask To the name correct?
- set user name
- say Your account is unfortunately disabled
- say Please talking to you (SignedName)
- say Due to technical issues I can't assist you for the moment. I apologize for the inconvenience
- say Because I can't assist you for the moment. I apologize for the inconvenience

Decision points include:

- if user is 'accountControlError'
- if user is 'none'
- if user is 'user'
- if user is 'user'
- if user is 'user'
- if user is 'user'

Process variables are shown on the right side of the ontology:

name	value

IGA

Current Spms	Parent Tag	Question / Action	Tags Specified
Account_password_demo	[subject Object]		
Account_password_demo	[subject Object]		

Brugt internt i 2016 til nulstilling af password m.m. Programmeret til at dække 15% af alle henvendelser og kunne håndtere 50% af disse (altså 7.5% samlet set).

SEB Aida

<http://www2.compute.dtu.dk/~tobo/SIRIkomSEB.mp4>

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 guider 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.



2. bølge til sprog: OpenAIs sprogmodel GPT-2, februar 2019

En sprogmodel som kan forudsige det næste ord i en sætning. Bruger neurale netværk (subsymbolsk).

Sentence:

I'd like a nice hot cup of

Options:

45.5% **coffee**
32.6% **tea**
2.2% **hot**
1.4% **jo**
1.0% **espresso**
0.8% **cocoa**
0.6% **water**
0.6% **chocolate**
0.5% **milk**
0.3% **ice**

Sentence:

I'd like a nice hot cup of skimmed

Options:

82.8% **milk**
2.7% **coffee**
1.5% **egg**
1.3% **water**
1.3% **tea**
0.7% **butter**
0.7% **-**
0.7% **lemon**
0.4% **sugar**
0.3% **cocoa**

(Screenshots fra den offentlige (mindre præcise) model, <https://gpt2.apps.allenai.org>)

Dronningens nytårstale—skrevet af GPT-2 sprogmodellen

Så er det nytårsaften. Vi har alle forskellige mål i livet. Men det danske nytår, der starter i dag, er en tid til at fejre, hvem vi er. Det er tid til at gøre noget andet. Det er tid til at give tilbage, at tage noget tilbage, at skabe noget nyt. Det er tid til at tænke på, hvad der er vigtigt i livet. Vi ønsker ikke bare at bevare det, vi allerede har.

Nogle vil hellere tilbringe tid med en, de elsker, end med en gruppe venner. Måske vil familien gerne spille et spil sammen eller lege tagfat. For et par år siden startede Danmark en sport, der er meget gammel, men stadig meget vigtig. Det er en lille turnering kaldet "Svenskning". Den finder sted hvert år. Her er tidsplanen: lørdag 15.-16. januar 20:00-19:00 og søndag 17.-18. marts 18:00-19:00. Jeg ønsker alle et godt nytår, og håber I forbliver sunde og raske! Hvis I har spørgsmål eller kommentarer, så kontakt mig endelig.

GUD BEVARE DANMARK

(Genereret af den offentlige version af OpenAI's sprogmodel GPT-2 på basis af tidligere af dronningens nytårstaler.)

Menneske vs maskine og fremtidens jobs

Computeres evner	komp. musik	spille skak	løse matematik opgaver	have paratviden (Watson)	smalltalk (sproglig int.)	forstå hinanden (social int.)
1950-65		✓	✓			
2010'erne	(✓)			✓	✗	✗



“AI will likely replace tasks rather than jobs in the near term, and will also create new kinds of jobs.”

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