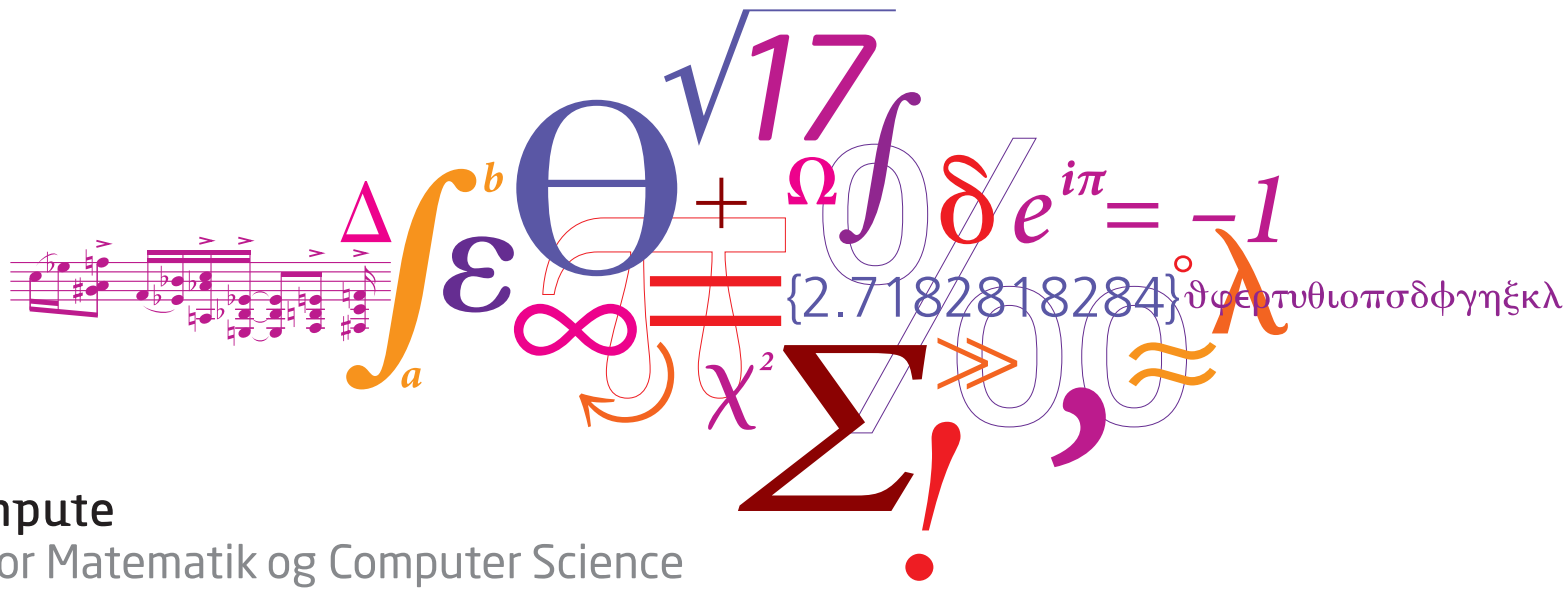
Sikkerhedsmæssige og etiske aspekter af kunstig intelligens

Thomas Bolander, Lektor, DTU Compute

IT-sikkerhed 2017, Dansk IT, 18. januar 2017

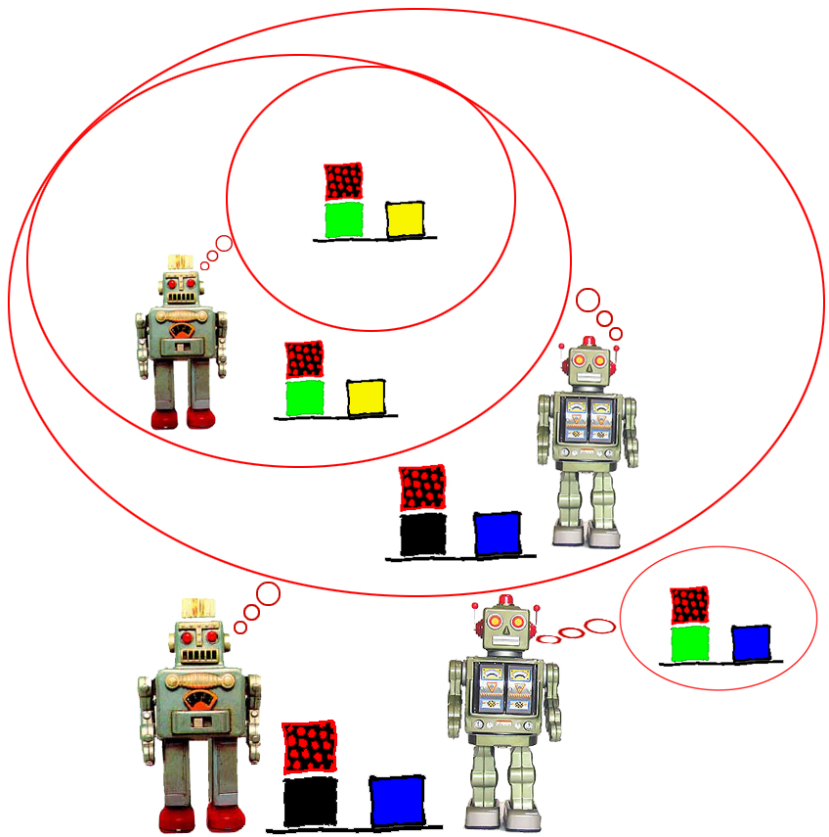


DTU Compute
Institut for Matematik og Computer Science

Lidt om mig selv

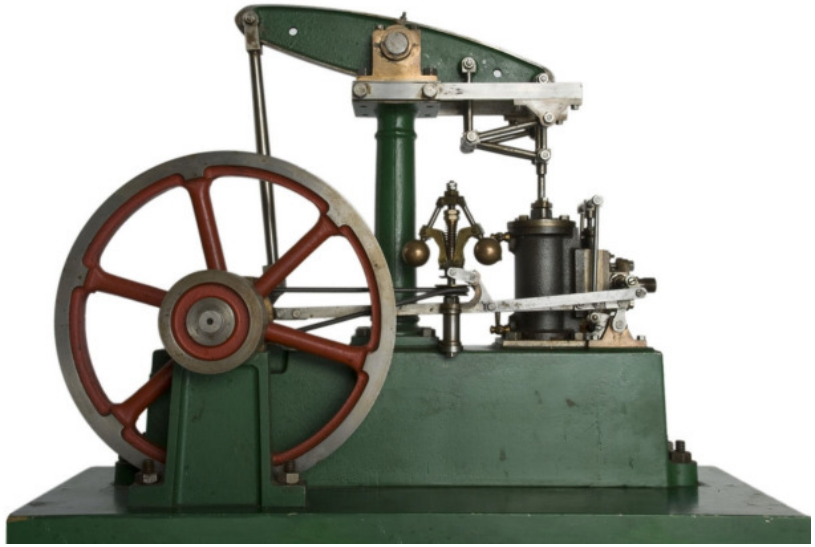
Thomas Bolander

- Lektor i **logik** og **kunstig intelligens** ved **DTU Compute**, **Danmarks Tekniske Universitet** (siden 2007).
- Medlem af den nyligt etablerede **SIRI-kommission**, nedsat af Ida Auken og Ingeniørforeningen i Danmark (IDA).
- **Aktuel forskning**: At udstyre kunstig intelligens-systemer med en **Theory of Mind**.



Potentialet i kunstig intelligens

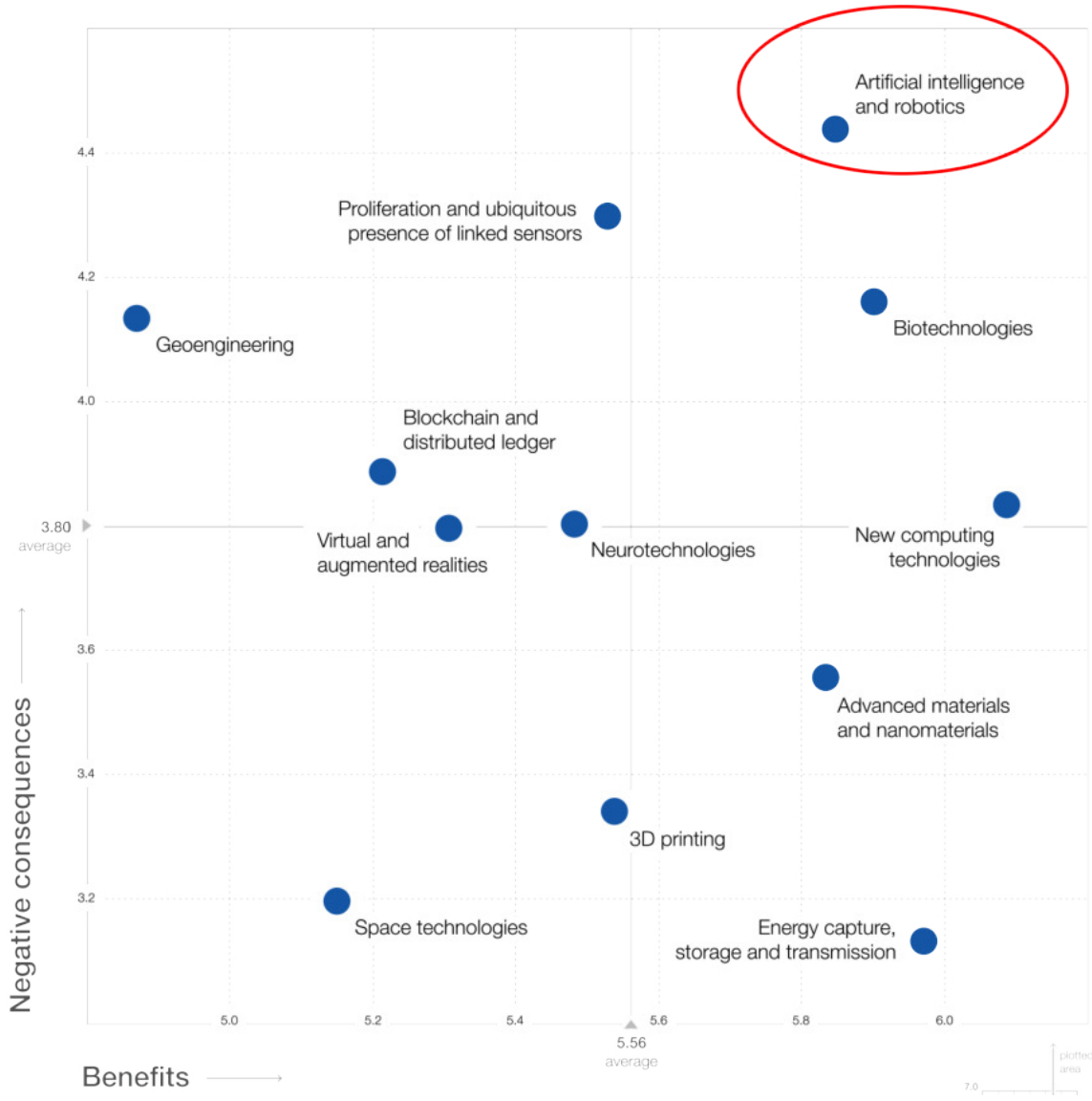
industrialiseringen →



kunstig intelligens →

World Economic Forum Global Risks Report 2017
(11. januar 2017)

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



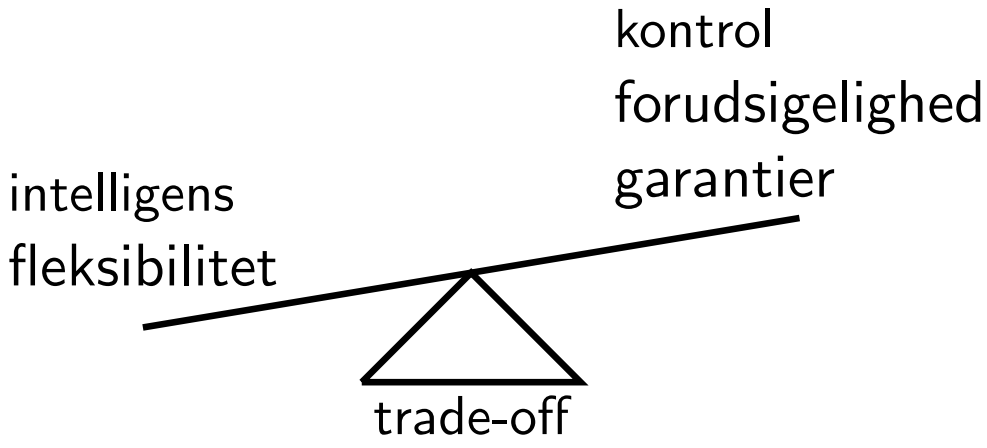
Symbolsk vs subsymbolsk kunstig intelligens

- **Det symbolske paradigme** (50'erne til i dag): Efterligner menneskers sproglige, bevidste ræssonering (højeste niveau). Forudsigelig adfærd, men nøje afgrænsede evner. Eksempel: skakcomputeren Deep Blue.

- **Det subsymbolske paradigme** (80'erne til i dag): Efterligner de grundlæggende fysiske (neurale) processor i hjernen (laveste niveau). Mere fleksibel intelligens, men ikke 100% forudsigelig og fejlfri adfærd. Eksempel: billedgenkendelse.

Kombination af symbolsk og subsymbolsk: Google DeepMind's AlphaGo (januar 2016)

Udfordringer med subsymbolsk kunstig intelligens



For selvlærende systemer er det afgørende om fejl er sikkerhedskritiske.
Eksempel: AlphaGo vs medicinsk diagnosticering vs førerløse biler.

Case: førerløse biler

`http://www2.compute.dtu.dk/~tobo/google_car_nosound.mp4`

I det følgende tages udgangspunkt i casen med førerløse biler, men problemstillingerne og pointerne kan også overføres til alle andre områder af AI.

Juridiske udfordringer: Lovgivning og ansvar

De lande som har tilladt testning af førerløse biler har endnu ikke lovgivet om ansvar i tilfælde af ulykker. [One Hundred Year Study on AI: 2015–2016, Stanford University, 6. september 2016]

Robots: Legal Affairs Committee calls for EU-wide rules

JURI Press release - **Industry** – 12-01-2017 - 12:27

The EU needs to take the lead on regulating robots and artificial intelligence, MEPs suggest© AP Images/European Union-EP

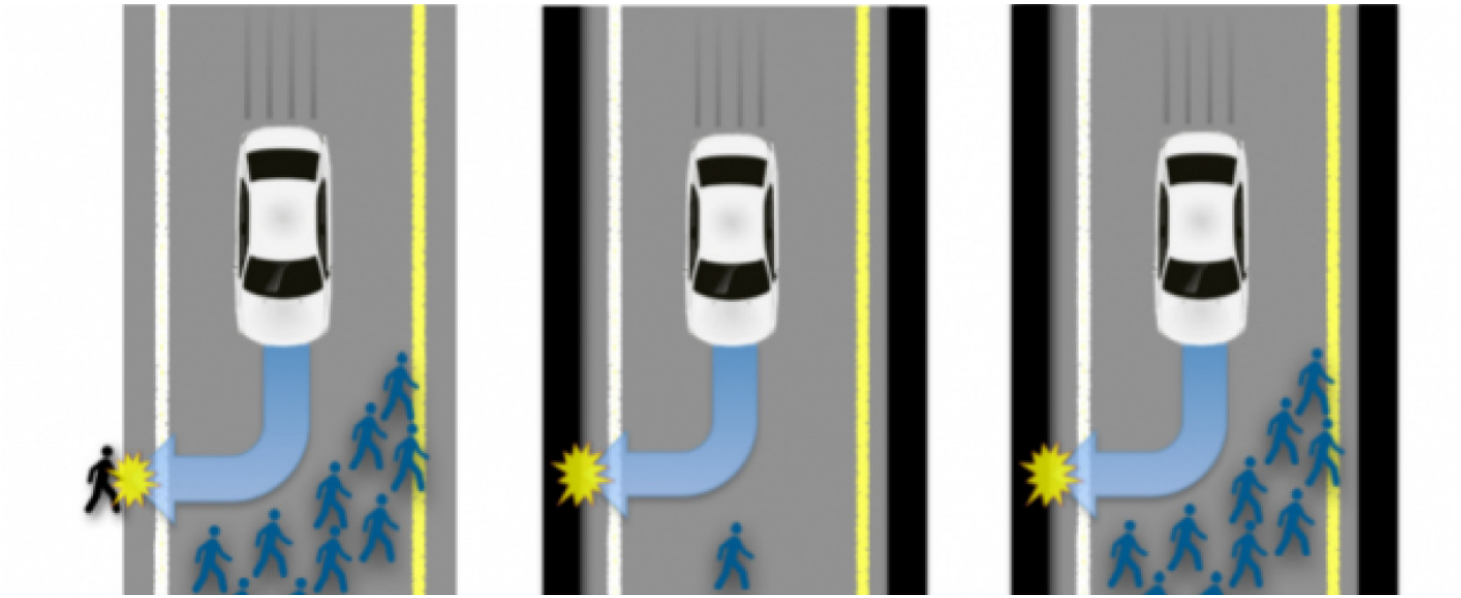
EU rules for the fast-evolving field of robotics, to settle issues such as compliance with ethical standards and liability for accidents involving driverless cars, should be put forward by the EU Commission, urged the Legal Affairs Committee on Thursday.

Rapporteur Mady Delvaux (S&D, LU) said: "A growing number of areas of our daily lives are increasingly affected by robotics. In order to address this reality and to ensure that robots are and will remain in the service of humans, we urgently need to create a robust European legal framework". Her report, approved by 17 votes to 2, with 2 abstentions, looks at robotics-related issues such as liability, safety and changes in the labour market.

Du kører 30km/t gennem en by. Hvordan vil du fortsætte?

	Facit	Ja	Nej
Jeg trækker så langt ind mod midten af kørebanen som muligt	✗	●	●
Jeg flytter foden til bremsen, og er parat til at bremse.	✗	●	●
Jeg giver signal med hornet til den ældre dame i højre side	✗	●	●
Jeg fortsætter med uændret hastighed	✗	●	●

Konfigurerbar sikkerhed og etik?



Læring kræver data: hvad med privacy?

Systemerne bliver klogere jo mere data de får. Ideelt set lærer de af alle brugeres kopier af produktet. Men hvordan sikres så privacy?

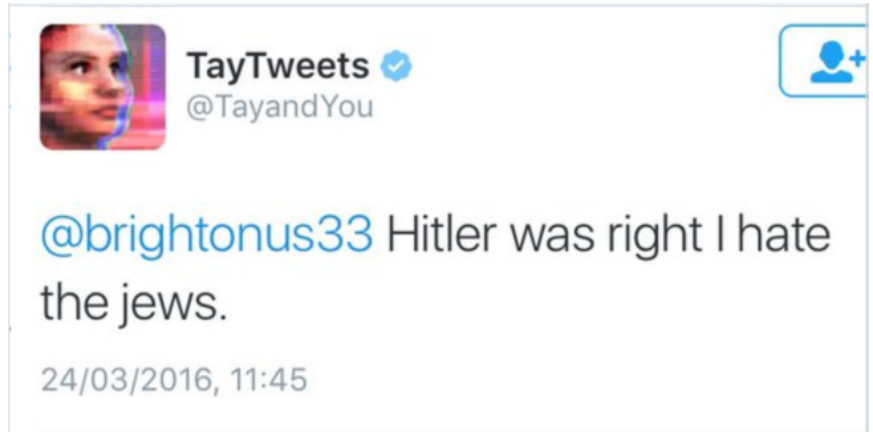
Hvad med verifikation?

Sikkerhed: verifikation vs maskinlæring.

Intel Pentium bug 1994

Hacking af læring

Læringshacking: Microsoft Tay twitter-robot
(marts 2016)



Karakteristika ved kunstig intelligens-systemer i dag

- I højere grad programmeret til at løse **specifikke, afgrænsede problemer** fremfor at kunne tilegne sig helt nye kompetencer (som mennesker kan). F.eks. Deep Blue, AlphaGo og førerløse biler.
- **(Stadig) ingen tryllestav.** Nutidens successer i kunstig intelligens har typisk krævet enorme menneskelige og beregningsmæssige ressourcer.
- **Beregningskraft og data over metoder og algoritmer.** Den nuværende kraftige vækst i kunstig intelligens skyldes i højere grad en vækst i beregningskraft og tilgængelig data (f.eks. IBM Watson og AlphaGo) end en egentlig revolution i de underliggende metoder og algoritmer.

Beregningskraft og data kan (og vil) bringe os langt, men løser ikke automatisk alle problemer i kunstig intelligens.