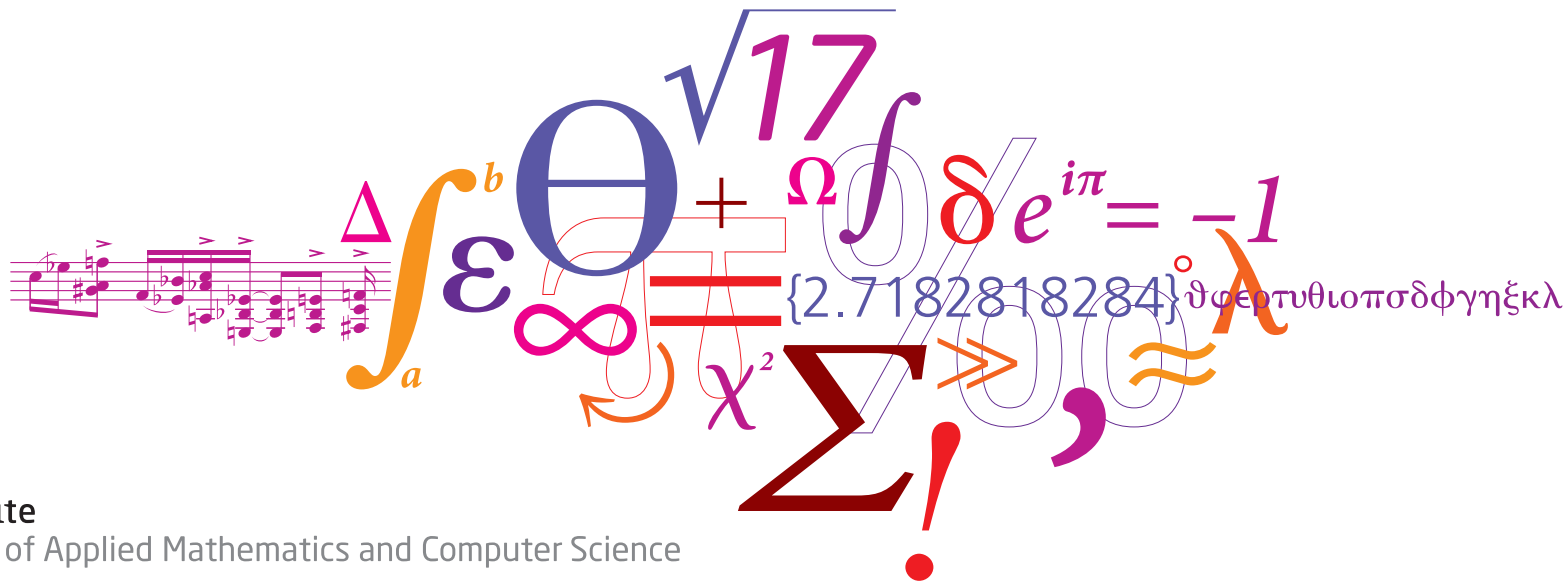
Human vs machine intelligence: How they differ and what this implies for our future society

Thomas Bolander, Associate Professor, DTU Compute

The Actor-Reality Construction, 25 October 2018

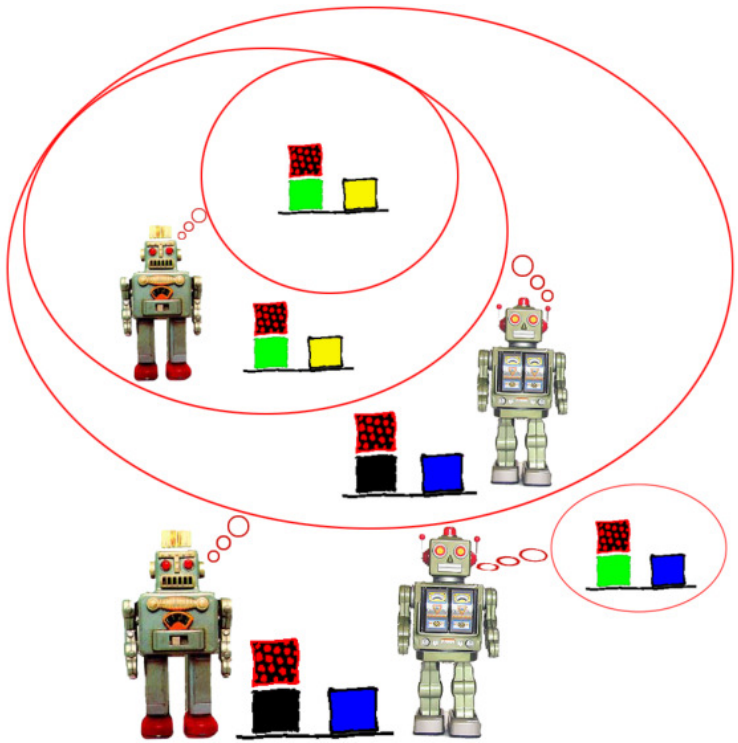


DTU Compute
Department of Applied Mathematics and Computer Science

A bit about myself

Thomas Bolander

- Associate professor in AI at **DTU Compute**, **Technical University of Denmark**.
- Member of the **SIRI commission**.
- **Current research**: Social aspects of AI. How to equip AI systems with a **Theory of Mind** (ToM).
- Co-organiser and scientific advisor for **Science & Cocktails** (`http://www.scienceandcocktails.org`).



What is artificial intelligence (AI)?

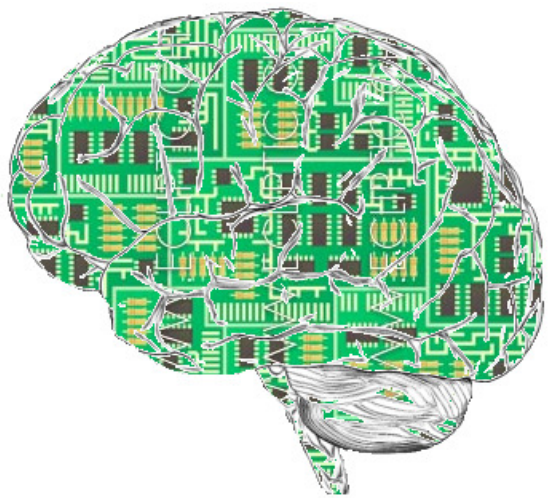
Definition by John McCarthy, the father of AI:

> *"Artificial intelligence is the* **science** *and* **engineering** *of making* **intelligent machines**, *especially* **intelligent computer programs**."
>
> *(McCarthy, 1956)*

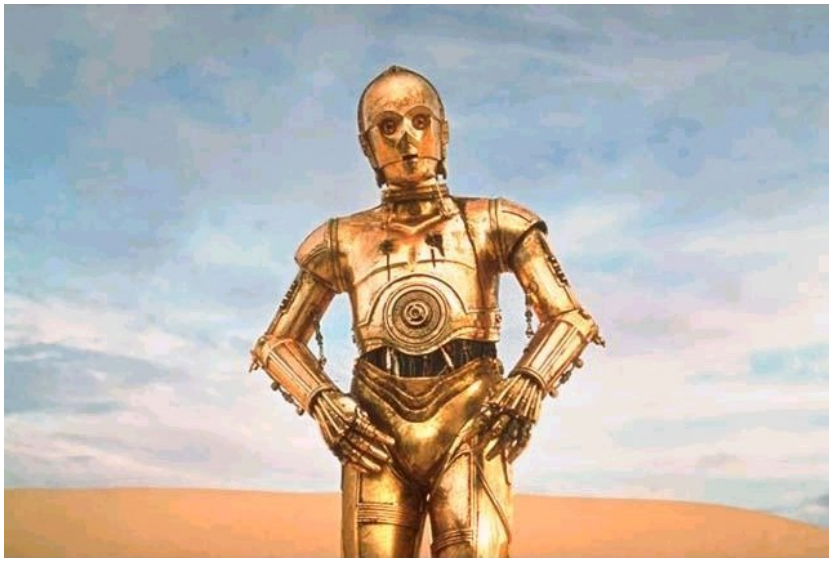
John McCarthy, 2006

Doesn't imply that they are intelligent in the same way as humans.

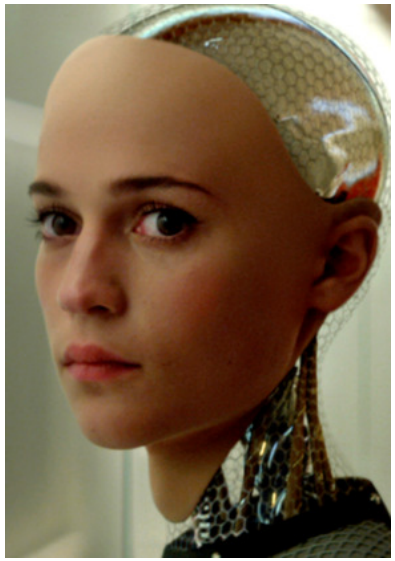
AI today is more different from human intelligence than anyone originally anticipated.

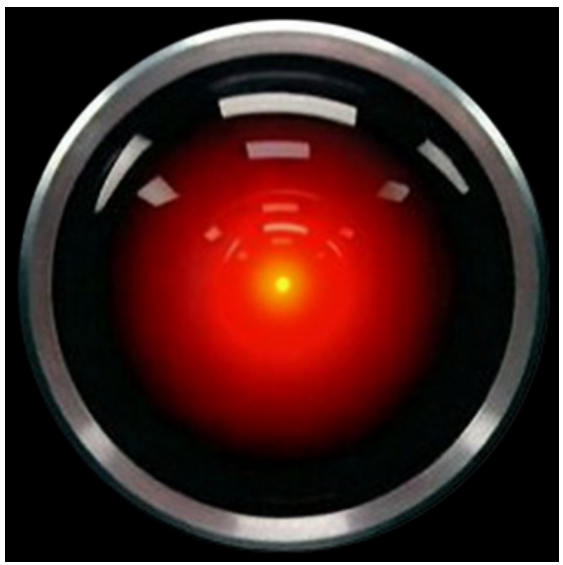


AI in sci-fi



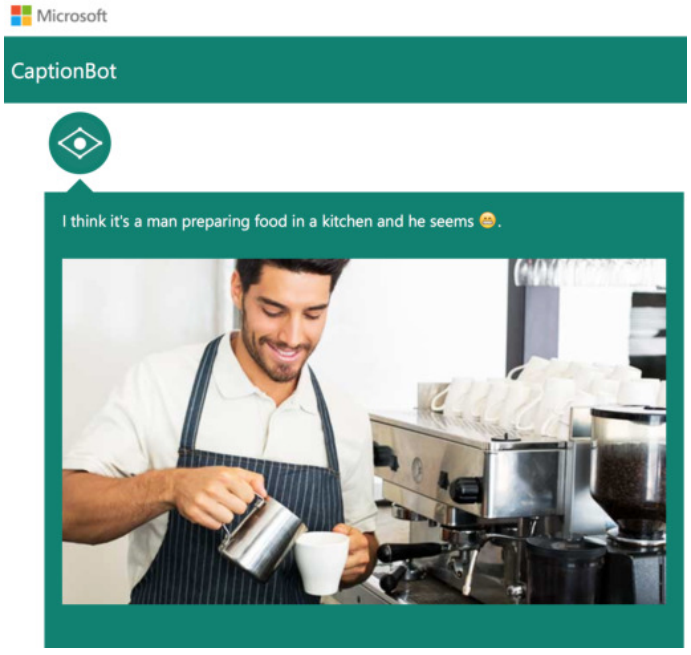

WALL·E






JOAQUIN PHOENIX
AMY ADAMS
ROONEY MARA
OLIVIA WILDE
AND SCARLETT JOHANSSON
her
A SPIKE JONZE LOVE STORY

AI in our everyday surroundings



CaptionBot image recognition

Siri on iPhone

Google driverless car

Characteristics of current AI

- **Current AI systems are specialised**: Solve well-defined, clearly delimited problems.
- **Rule of thumb**: The more well-defined and clearly delimited a task is, the easier it is make an AI system that can do it.

IBM Watson (2011): Jeopardy world champion

- 200 million pages of text in memory.
- Processes 1.000.000 books per second!

Problem solving is a combination of:

1. Ability to extract **information from data** (intuition, abstraction, conceptualisation).
2. Ability to **process data quickly** (search).

Often a deficiency in 1 can be **compensated** by a dramatic increase in 2.

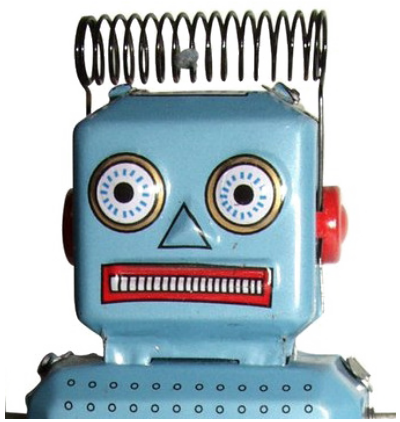
Human-machine dualism

Human

Flexible intelligence, abstract thinking, good at conceptualising the world.

Information from data 

Process data quickly 



Machine

Good at clearly delimited and well-structured problems.

Information from data 

Process data quickly 

Symbolic vs sub-symbolic AI

The symbolic paradigm (1950–): Simulates human symbolic, conscious reasoning. Search, planning, logical reasoning. **Ex**: intelligent personal assistants.


robust, predictable, explainable


strictly delimited abilities


flexible, learning


never 100% predictable/error-free

The sub-symbolic paradigm (1980–): Simulates the fundamental physical (neural) processes in the brain. Artificial neural networks. **Ex**: image recognition.

symbolic

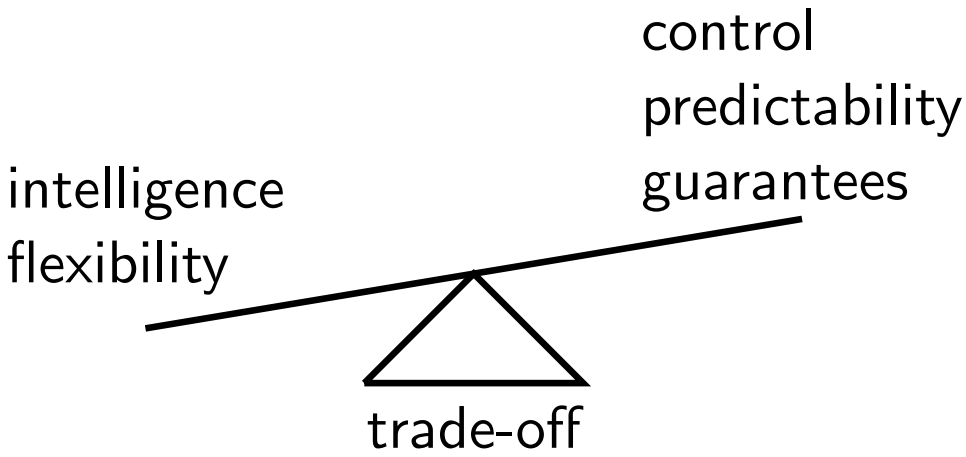
sub-symbolic

Challenges in subsymbolic artificial intelligence

Tesla crash, June 2016

Uber Volvo accident, March 2018

Subsymbolic techniques like neural networks can never become 100% precise in their classifications.



Trust from low probability of mistakes?

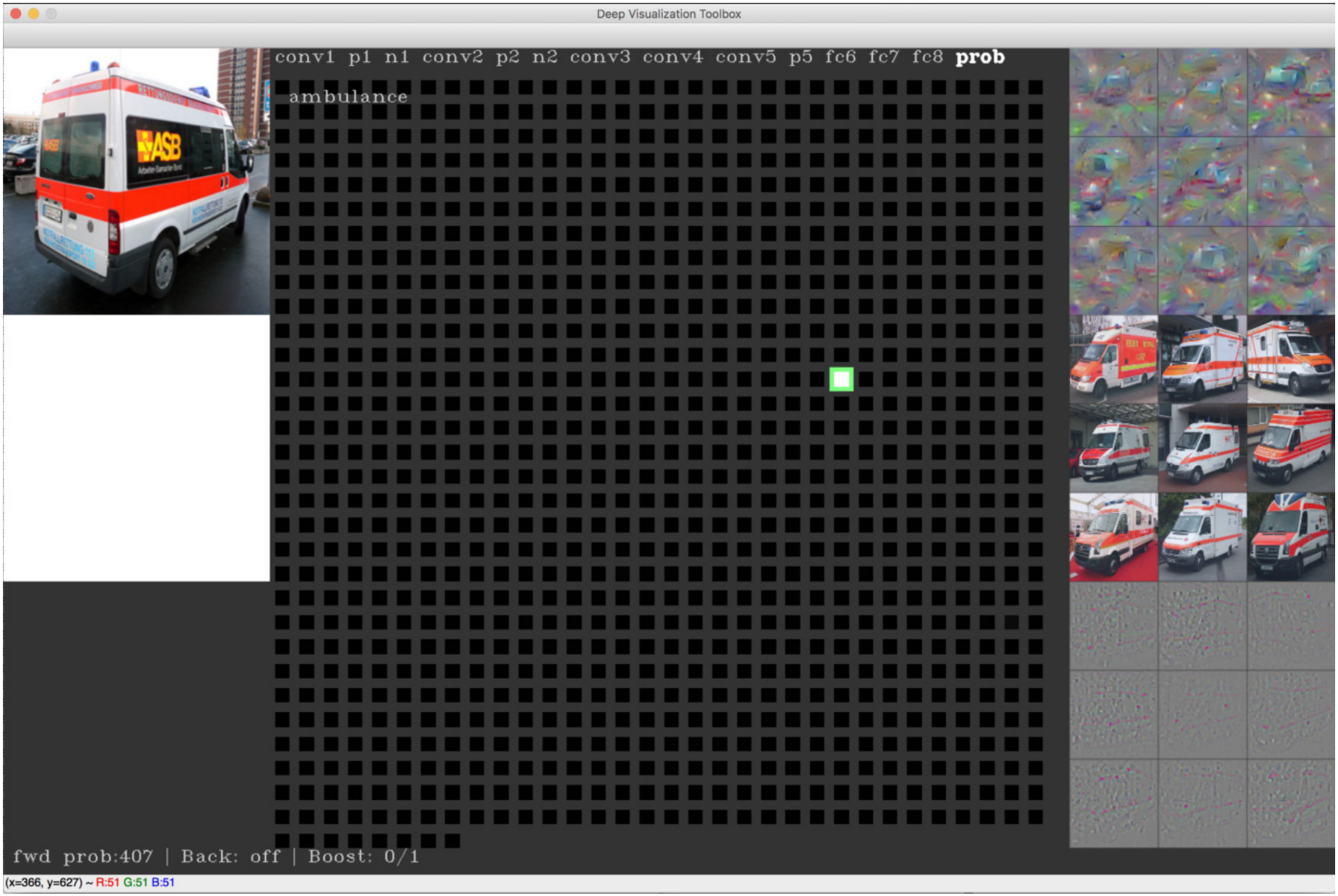
Is it sufficient that *the agent is almost always right*?

No:

> *For example, we cannot argue that a pedestrian detector is safe simply because it performs well on a large data set, because that data set may well omit important, but rare, phenomena (for example, people mounting bicycles). We wouldn't want our automated driver to run over a pedestrian who happened to do something unusual.*

(Russell & Norvig: Artificial Intelligence—A Modern Approach, 3ed, 2010.)

Demo of deep neural networks



Trust and explainability: The relationship between humans and machines

When do we trust the **decisions** of AI:

1. When they **never make mistakes**?
2. When they **almost never make mistakes**?
3. When they **most often don't make mistakes**, but when they do, they have an acceptable and **explainable reason**.

They [AI systems] should be designed to enable people to **understand AI systems successfully**, *participate in their use, and* **build their trust**.

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**.

(One Hundred Year Study on AI: 2015–2016, Stanford University, Sep 2016)

Symbolic vs sub-symb. AI: explicit vs implicit models

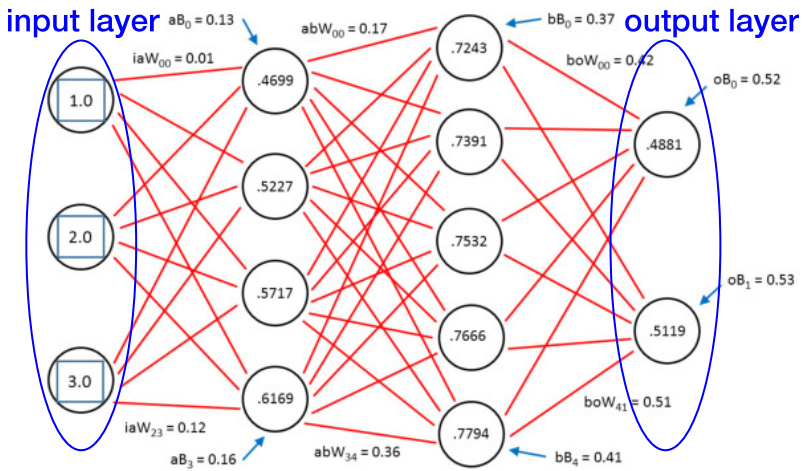
Explicit/symbolic

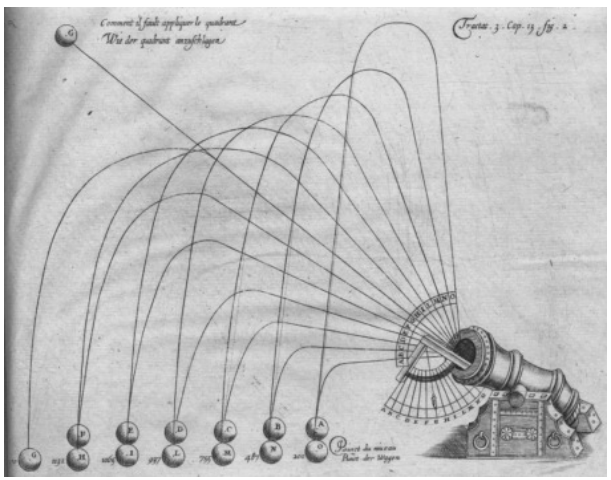
$$a_x = \frac{-kv_x}{m} = \frac{dv_x}{dt} \text{ (1)},$$

and

$$a_y = \frac{1}{m}(-kv_y - mg) = \frac{-kv_y}{m} - g = \frac{dv_y}{dt} \text{ (2)}$$

Implicit/subsymbolic





From raw data to symbolic representations

Subsymbolic representation (data)

cat[1]

NOUN (plural **cats**, plural **cats**)

1 A small domesticated carnivorous mammal with soft fur, a short snout, and retractable claws. It is widely kept as a pet or for catching mice, and many breeds have been developed.

Symbolic representation

- **Subsymbolic AI**: input is raw data (subsymbolic), output is subsymbolic (implicit model).
- **Symbolic AI**: Input is symbolic, output is symbolic (explicit model).

What we really need for **explainability**: input is raw data, output is explicit model (symbolic). Requires combining symb. and subsymb. AI.

Regaining trust: explainable AI

- Trust in AI systems is at risk when systems are neither 100% **robust**, nor **explainable** (by themselves or from the outside).
- In lack of 100% robustness, we need more **transparent** and **explainable** AI.
- **Subsymbolic AI** (e.g. neural networks) is naturally opaque.
- **Symbolic AI** (e.g. manually hand-crafted rule-based systems) is naturally transparent, but difficult to craft.
- Best current bet is to **combine**: The output of learning is rules and explicit models that can be inspected, understood and modified by humans.

The Big Data mantra of "what, not why" (e.g. Mayer-Schonberger & Cukier: Big Data—A Revolution that will Transform how we Live, Work and Think) is challenged when decisions are made by algorithms, and the people affected want an explanation.

The 3 hardest problems in AI

Social intelligence: The ability to understand others and the social context effectively and thus to interact with other agents successfully.

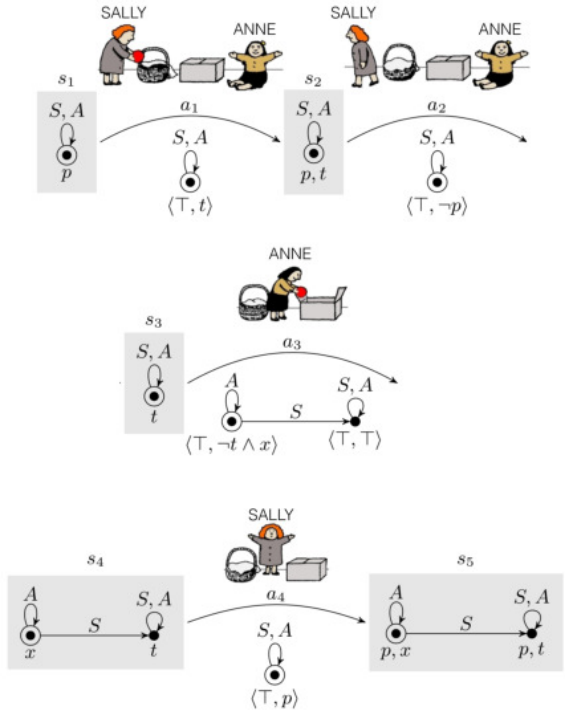
Carl Frey, 20 April 2017
Kolding, Denmark

Toby Walsh, 18 March 2017
Science & Cocktails, Copenhagen

Both have **social intelligence** among the 3 human cognitive abilities that are hardest to simulate by computers and robots.

My current research in robots with social intelligence

subsymbolic



symbolic

Solving cognitive tasks: **false-belief tasks** of arbitrary order. Humans can solve first-order at age 4, second-order at age 10, third-order at age 20.

- **Sub-symbolic** (perception): face/object recognition, skeleton tracking, speech-to-text.
- **Symbolic** (higher cognition): planning, intentions, logical reasoning, perspective-taking.

The impact of artificial intelligence (AI) on the human competences of the future

Required human competences of the future:

1. Competences in **seeing the potential** and **selecting** tasks to be automatised. Requires: *Overall understanding of AI*.
2. Competences in **implementing** AI techniques for 1. Requires: *Technical understanding of AI*.
3. Competences to **operate** and **collaborate** with AI systems. Requires: *Overall understanding of AI*.
4. Competences in areas **that can not be automatised**. Requires: *?*.

Most important, and affecting most people, is 3 and 4.

Instagram image recognition (2015)

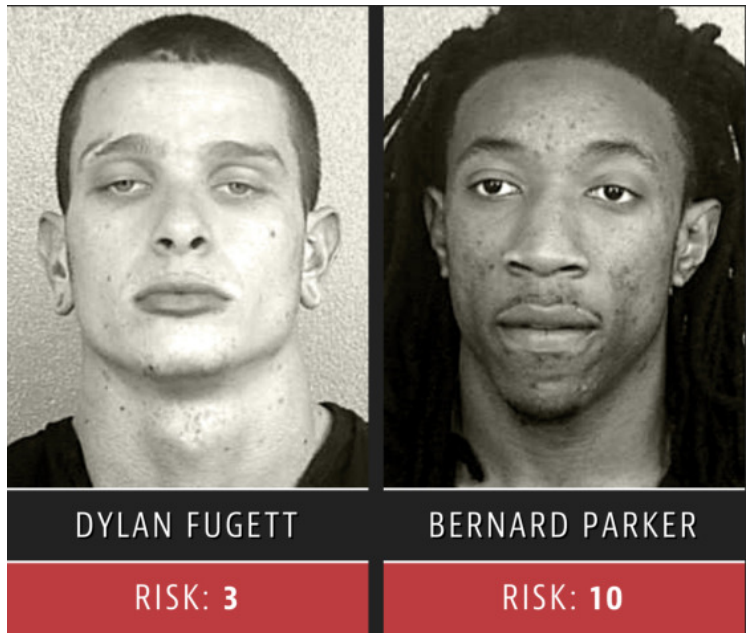
'Your account has been disabled for not following the Instagram Community Guidelines, and we won't be able to reactivate it.

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)

AI for risk assessment (2016)

Great potential in AI for **classification** (credit scoring, filtering job applications, etc.). But the systems only build an implicit model of correlations in data (**not** causal relationships), and can easily become biased.



	WHITE	AFRICAN AMERICAN
Labeled Higher Risk, But Didn't Re-Offend	23.5%	44.9%
Labeled Lower Risk, Yet Did Re-Offend	47.7%	28.0%

(Angwin et al.: Machine Bias, ProPublica, 23 May 2016)

Automatisation vs human competences

Australian study (Reeson et al., 2016) of the impact of digitalisation/automatisation on the required competences of skilled and technical workers concluded:

> *Across the areas with the highest growth in new jobs in the period 2011–2015, competences related to* **communication** *was rated highest. The interviews furthermore suggested that competences like* **social empathy** *and the ability to* **critically evaluate** *digital data sources (text, sound, images).*

Kunstig intelligens—Morgendagens job og samfund, SIRI-kommissionen 2016. (my highlightings)

In a separate report from the SIRI commission, we concluded: To utilise the potential of digitalisation/automatisation, the most import thing is to make the employees feel safe, not fearing the technology, not fearing their jobs. Not about **skills**, but about **attitudes**.

Human + machine: AI as teaching assistant (Georgia Tech, 2016)

The chatbot Jill based on IBM Watson becomes teaching assistant.

Student	Should we be aiming for 1000 words or 2000 words? [...]
Jill	There isn't a word limit, but we will grade on both depth and succinctness. It's important to explain your design in enough detail [...]
Student	Jill can you please elaborate on "it's important to explain your design in enough detail". What kind of design are you referring to?
Human TA	I think Jill is using "design" as a catch-all statement. For the midterm, it refers to [...]

Many answers are simply excerpts from assignments and other course material. But still has a great value.

The number of questions increase with the number of students, but not the number of **different** questions.

Humans vs machines and the jobs of the future

	comp. music	play chess	do math	trivia know-ledge	**smalltalk (linguistic)**	**understand others (social)**
1950s+1960s		✓	✓			
2010s	(✓)			✓	✗	✗

"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 Sep 2016)