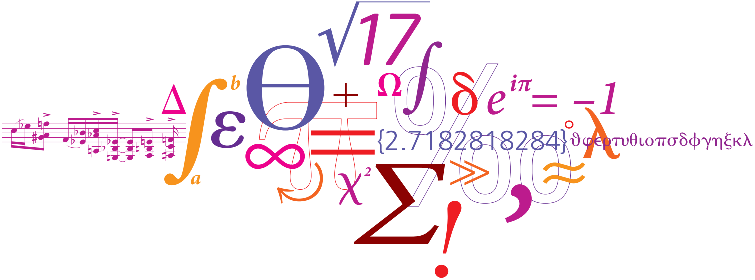
What do we lose when machines take the decisions?

Thomas Bolander, DTU Compute, Technical University of Denmark



DTU Compute
Department of Applied Mathematics and Computer Science

A bit about myself

Thomas Bolander

- Professor in logic and AI at **DTU Compute**, **Technical University of Denmark**.
- **Current research**: Social aspects of AI. To equip AI systems with a *Theory of Mind* (ToM).
- Member of several commissions and think tanks concerned with the ethical and societal aspects of AI.
- Co-organiser and scientific advisor for *Science & Cocktails* (`http://www.scienceandcocktails.org`).

What do we lose when machines take the decisions?

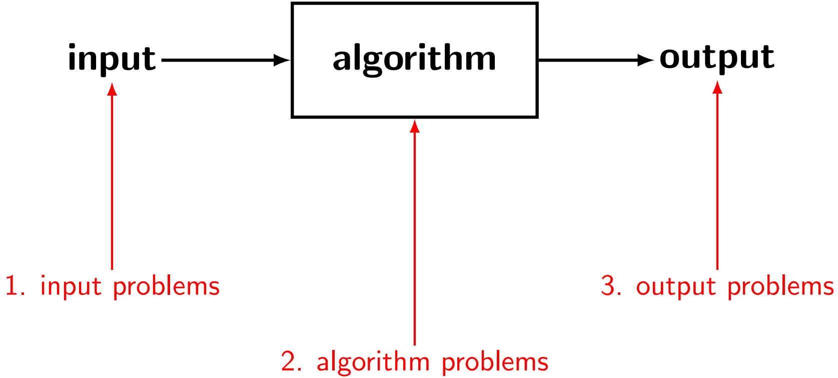
What do we lose when machines take decisions (or do categorisations, or make predictions, or do rankings)?

Any difference comes down to **differences in human intelligence and machine intelligence**. These differences are defined by limitations in the current level of AI.

Reasonable principles:

1. Technology should enhance transparency, fairness and explainability, not diminish them.
2. When a task is (partly) automatised, require at least the same level of problem solving as before automatisation—on all parameters.

Where are the problems in AI—compared to human intelligence?



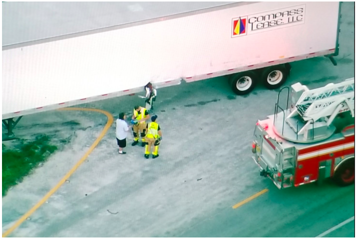
1. Input problems: Too little data...

Missing dimensions. E.g. language models.

> *A recent study performed by the Department of Psychology examined the impact of the Digital-Human Interaction Model on the evaluation of performance in performance management and organizational decision-making as well as the assessment of the psychological health of employees using the Digital-Human Interaction Model (Gorzalescu, 2010).*

Missing data points

Tesla crash, June 2016



Tesla crash, March 2019

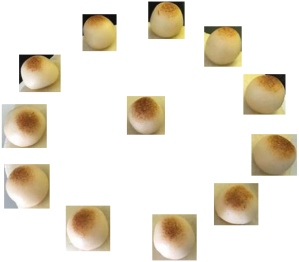
1. Input problems: ...or too much data

The essential problem is this:

- More context in data ⇒ less generalisable data.
- More dimensions in data ⇒ less generalisable data.

Solution in AI: less context, fewer dimensions!

This goes against what we want: Fair and valid assessments require detailed data, preserving contextual information.



Prediction of cinema visits: Too much or too little data?

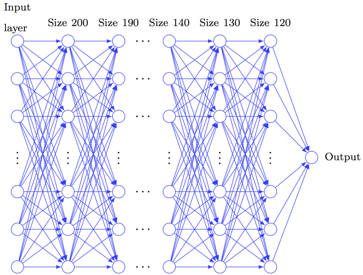
Collaboration between the IT-University of Copenhagen (ITU) and Nordisk Film (2017).

Method: Deep neural networks.

Input: Genre, budget, country, prequels, rating, cast, length, Google Trends, Twitter, Wikipedia, prerelease, competition, year, weather, reviews, source material, amount of marketing, preorders.

Output: Number in the range 1–9, categories of how many tickets sold.

Result: Predictions of human experts are still far better.



NORDISK FILM

2. Algorithm problems

How do most current AI algorithms solve problems?

1. **Pattern recognition (primarily deep neural networks)**. Pattern recognition can be used for classification, but goal and purpose is separated from the classification task—and so is the learning.
2. **Reinforcement learning**. Goal is integrated in the algorithm, but learning is not generalisable to other goals.

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

Doing the right thing is not **just** about being shown what the right thing is (1) or learning it through trial-and-error (2)...

Human problem solving

`http://www2.compute.dtu.dk/~tobo/children_cabinet.mpg`

The child (18 months) is *not* given any instructions. [Warneken 2006]

Required: 1) goal recognition; 2) perspective-taking (ToM); 3) multi-agent planning.

What is needed for organisational decision making?

Quotes about performance management (my highlightings):

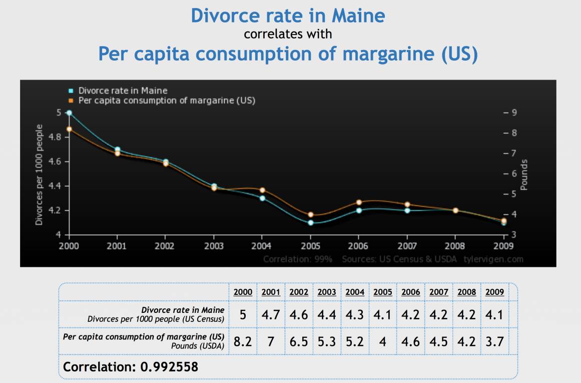
> *They assert that substantial gains in performance are more likely to be achieved by management* **understanding how employees perceive the world** *and then encouraging and implementing changes that make sense to* **employees' worldview***.*

(Wikipedia, referring to the US Office of Personnel Management)

> *Ultimately, every performance management system should ensure the achievement of overall organisational goals and ambitions* **while aligning them with employee goals***.*

(`https://www.clearreview.com/why-performance-management-important/`)

Statistical correlations vs causal relationships



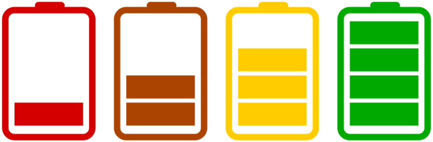
Divorce rate in Maine

correlates with

Per capita consumption of margarine (US)

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Divorce rate in Maine *Divorces per 1000 people (US Census)*	5	4.7	4.6	4.4	4.3	4.1	4.2	4.2	4.2	4.1
Per capita consumption of margarine (US) *Pounds (USDA)*	8.2	7	6.5	5.3	5.2	4	4.6	4.5	4.2	3.7

Correlation: 0.992558



battery level:

credit score: 1 3 7 10

3. Output problems

- Binary decisions
- No explanations.
- No dialogue.

Example: The National Danish AI Strategy (The Danish Government, March 2019)

Signature project: targeted employment efforts

> *With the help of artificial intelligence, it will potentially be possible to reduce the period of unemployment. By analysing patterns in historical data on successful efforts, the caseworker will have a better opportunity to target employment efforts to the individual citizen.*

1. **Input problems**. E.g. missing relevant data about individuals normally obtained through human dialogue.
2. **Algorithm problems**. Patterns, not causal relationships. No complex reasoning about the goal of the effort, and the possible ways to get there.
3. **Output problems**. Only decision, no dialogue, no explanation.

Combining symbolic and 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.

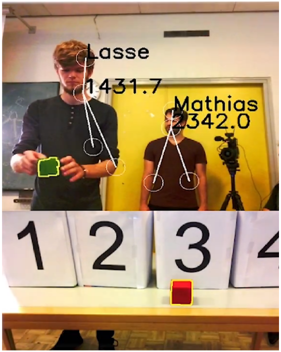
symbolic

↓

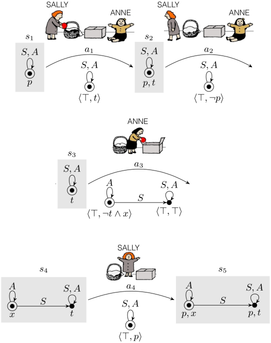
↑

sub-symbolic

Combining paradigms: my research on social robots



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): image recognition, object recognition, skeleton tracking, speech to text.
- **Symbolic** (higher cognition): planning, intentions, logical reasoning, perspective-taking.