Chancen und Grenzen von Autonomous Driving in Mobilitätsszenarien

Dr. Gabriel Seiberth

November 3, 2022

The promise…

March 2015

Tesla CEO Elon Musk and Nvidia CEO Jen-Hsun Huang declare self-driving cars "solved"

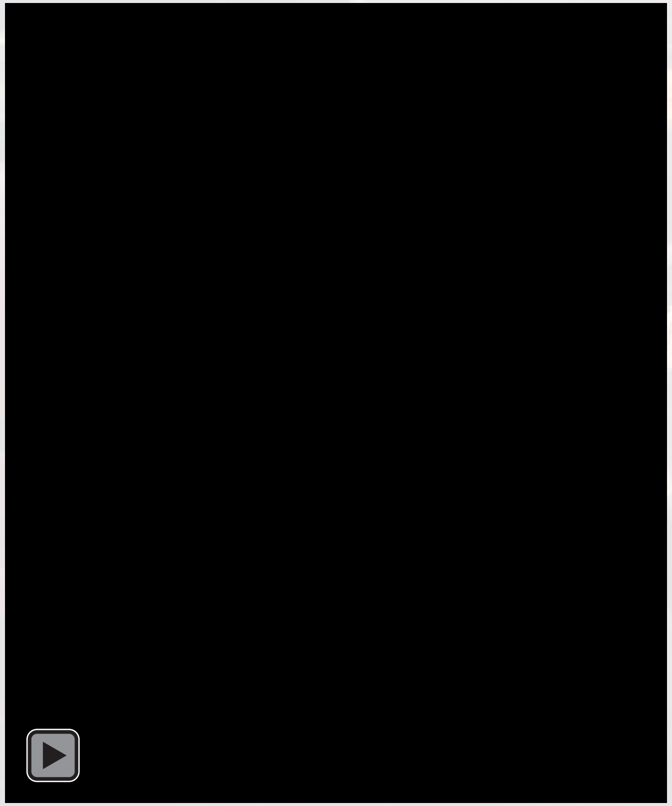
Fortune (2015)

The reality…

2020

Tesla's latest „Full Self-Driving" update recognizes supermarket flags in the wind as switching traffic lights

TESLA PERCEPTION MISCLASSIFICATION



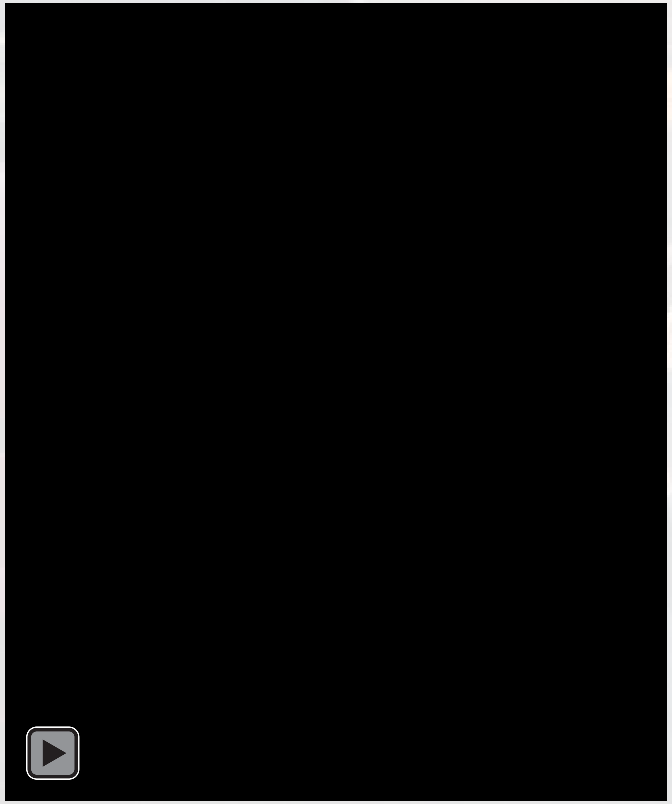
Evgeny Becker, LinkedIn (2020)

The reality...

2021

*Tesla's latest „Full Self-Driving" update recognizes **traffic lights mounted on trailer while going 130 km/h on the highway***

TESLA PERCEPTION MISCLASSIFICATION



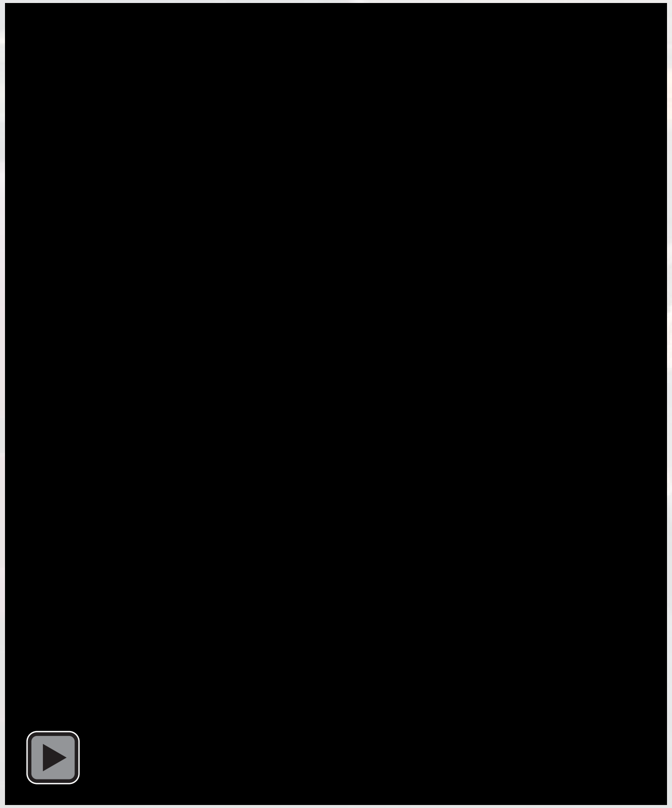
Twitter (2021)

The reality...

2021

Tesla's „Full Self-Driving" system slowing down vehicle after mistaking the moon for a yellow traffic light

TESLA PERCEPTION MISCLASSIFICATION



NDTV (2021)

The reality…

2022

Tesla's latest „Full Self-Driving" update recognizes horse-drawn coach as pedestrian, vehicle or truck.

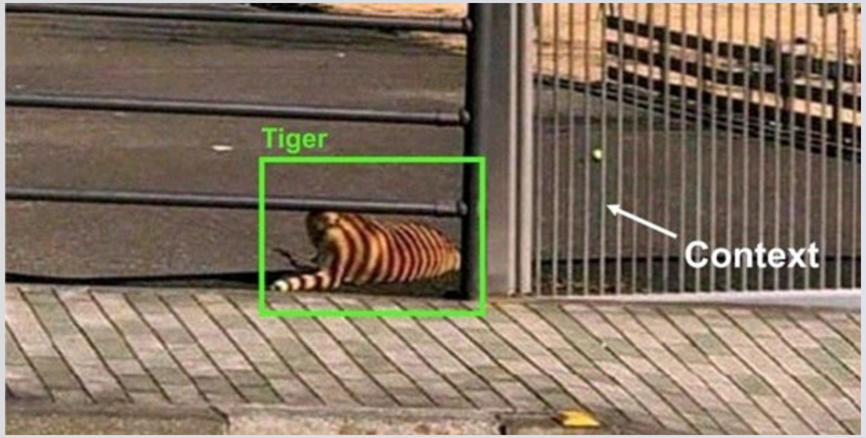
TESLA PERCEPTION MISCLASSIFICATION

LinkedIn (2022)

What do all these cases have in common?

They demonstrate the barrier of meaning
in task-based AI –
and that cannot just be
engineered away

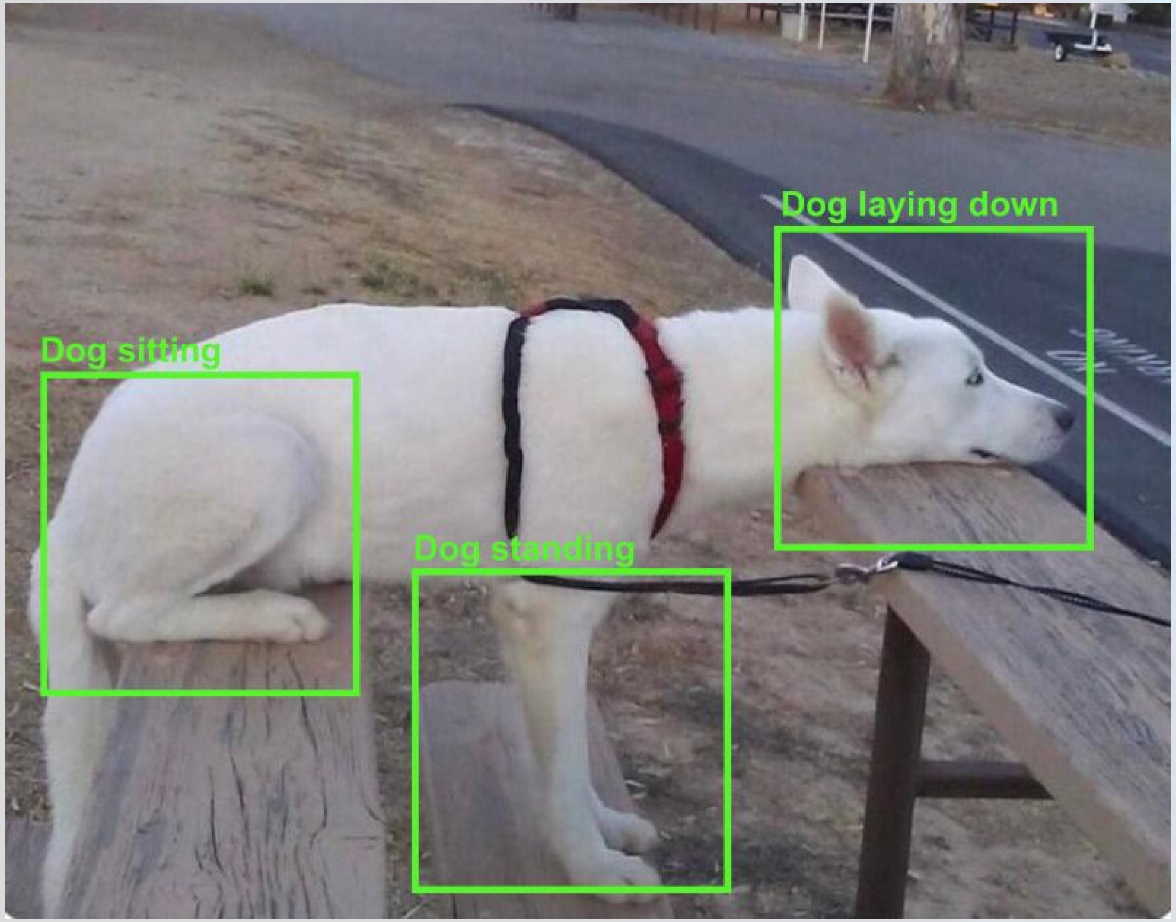
Barrier of meaning 1/3



Twitter (2022)

Key challenge: **Context**

Barrier of meaning 2/3



Ralph Aboujaoude Diaz (2022)

Key challenge:
Ambiguity

Barrier of meaning 3/3

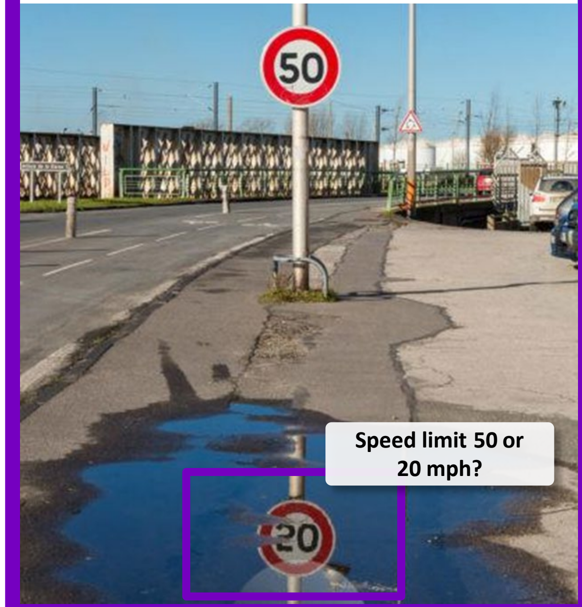
Bevan Hurley (2022)

Key challenge:
Compositionality

Barrier of meaning in practice: AI systems are easily fooled – unwillingly and – even worse – willfully

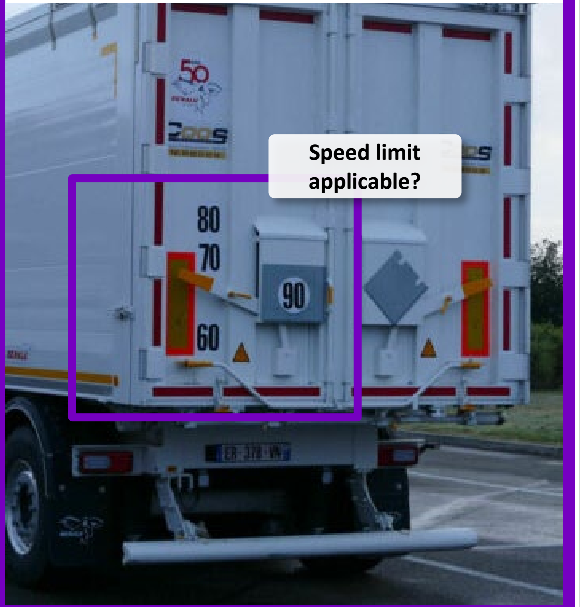
Unwillingly

TRAFFIC SIGN REFLECTIONS



Julia Pfeifer (2016)

SPEED LIMITS ON TRUCKS



Motortalk (2018)

Willfully

45 MP/H SPEED LIMIT

Cornell University (2018)

ADVERSARIAL PATCHES

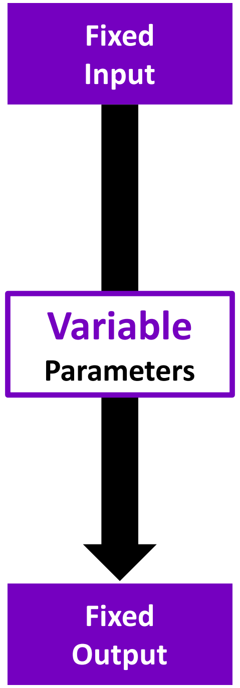
Cornell University (2020)

Statistical curve-fitters perceive the world in a completely different ways as humans – which **opens the door for profound manipulation**

What is happening?

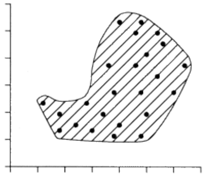
Supervised Learning is trapped in the training data set – Neural Networks struggle with transfer of knowledge

Supervised Learners
(e.g. Neural Networks)



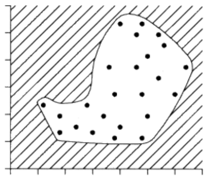
Due to **fixed inputs and outputs,** the learning process within supervised learning is **trapped in a pre-defined training set!**

Supervised learning just has the ability of
interpolation



Deep neural networks are **doing well with classification within known training space** (within known classes) and with **similarity between training and testing data**, even with examples which are never seen before

…but no ability of **extrapolation**



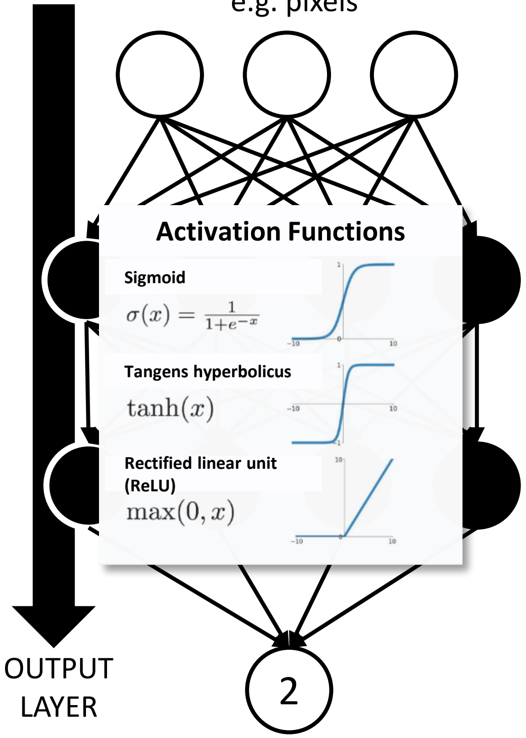
Supervised learners **cannot extrapolate** – that means they **cannot handle new situations** (outside their training space) **by adapting previously learned inputs**
Melanie Mitchell (2020)

> Then, we run a machine-learning algorithm against the training set to construct a model, a block box, that **predicts what we already know**.
>
> *Meredith Broussard (2018)*

Ultimately, learning in Neural Networks is curve-fitting

“Learning“ in Deep Learning = Minimizing a Cost Function

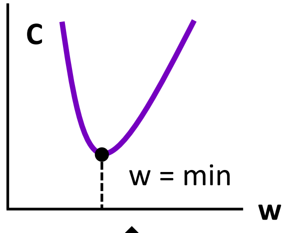
INPUT LAYER



OUTPUT LAYER

e.g. numbers

Backpropagation =
Algorithm for finding that gradient



C $(w_0, w_1, \ldots, w_x+b)$ = **Cost**

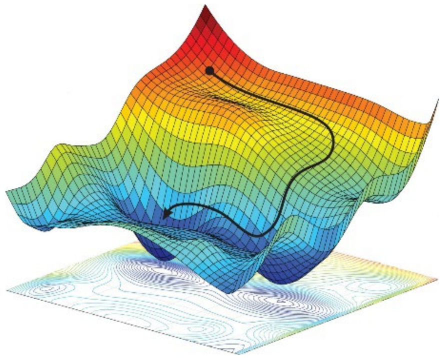
Real Output

Desired Output

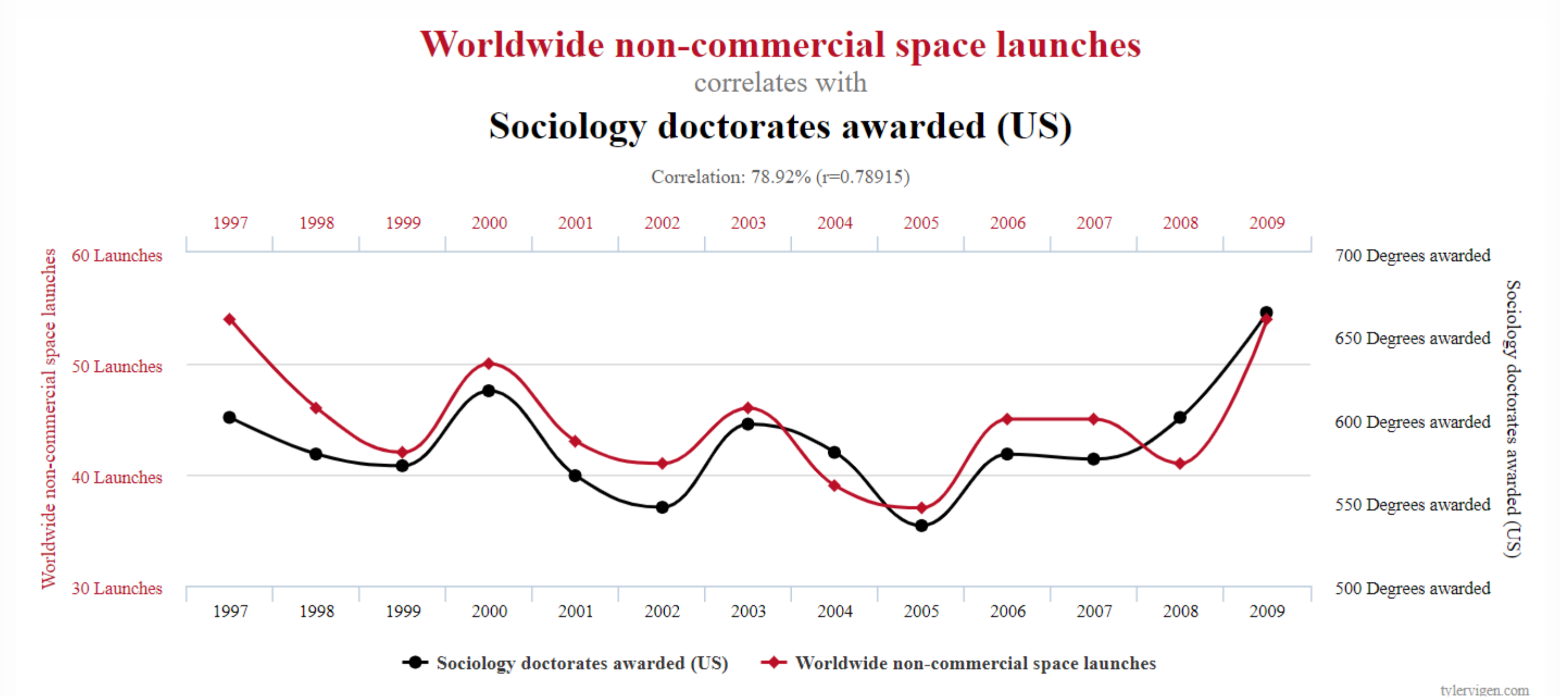
Gradient Descent =
Find the steepest descent (negative gradient) to decrease the cost function (**C**) by adjusting weights (**w**) and biases

Cost Function = calculates the **Cost** within the network by comparing real with desired output

Cost =
Differences between real output and desired output



Neural networks use statistics, they are model blind and cannot distinguish correlation from causality



Correlation ≠ Causation

The Big Data Paradox:

“Too much information tends to behave like **very little information”**

Calude and Longo 2017

Neural networks are **collectively impressive, individually unreliable.**

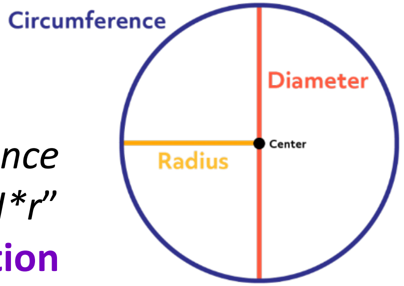
Not all relevant knowledge stems from observation

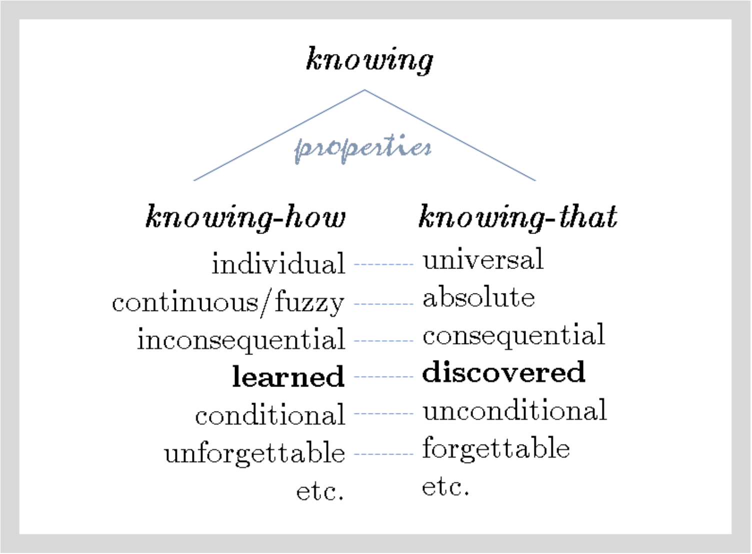
example

"***knowing how*** *to play guitar*"
→ **learned by experience (observation)**

vs

"***knowing that*** *the circumference of a circle is 2*PI*r*"
→ **learned by "being told" / deduction**





Commonsense knowledge cannot be learned, but is **based on "previously universally valid cognitive templates*** that we don't even learn at all. [...]"

Sources: Medium (2022)

Unlike AI, humans are good at knowledge transfer – they can use commonsense, logic, analogies and inference

No transfer

X

CHESS

GO

Similarities:

- Two players
- Rule-based (static rules)
- Board game
- Perfect Information
- Two-colored playing field
- Finite number of moves
- Finite number of fields
- Well-defined moves per game piece
- Reliable prediction possible (‘closed world’)

“

DeepMind AlphaZero’s algorithm can be trained to play both Go and chess, but not both at once.
Retraining a model’s connections and responses so that it can win at chess resets any previous experience it had of Go.
If you think about it from the perspective of a human, this is kind of ridiculous.
People don’t forget what they’ve learnt so easily.

”

Douglas Heaven (2019)

Basically, **common sense is a model of the world**, which is **commonly shared, partially innate** and **updated continuously** over time – we **cannot describe** it and **take it for granted**.

Despite generative approaches, models still rely on gradient-based learning, with no common sense

**DALL-E EXAMPLE:
GENERATED IMAGE BASED ON TEXT**

"A coffee cup with many holes"

Irina Blok (2022)

No comprehension & reasoning due to missing model of the world

Correct depiction of example would require concept of "container".

Human ability of **reasoning based on radically incomplete information requires complex rule & knowledge-based system**, e.g. for understanding concept of **"container"**

Ernest Davis, Gary Marcus, et al. (2017)

AI systems are **missing a model of the world, common sense**, **reasoning and ability to understand complex phenomena**, like time, space and physics.

What does this mean for mobility scenarios?

The limitations of Deep Supervised Learning lead to three definite prerequisites of AD systems

AD systems require limitations, specified within their Operational Design Domain (ODD)

1
Environment Restriction

Specific Operational Conditions:
› Weather
› Geography
› Driving speed
› Roadways

2
Auxiliary Technologies

Supporting Technologies:
› Lidar
› HD maps
› V2X connectivity

3
Human Supervision

Human as central part of the equation:
› Safety driver
› Remote teleoperator
› Remote Supervision

“ODD defines the limits within which the driving automation system is designed to operate, and as such, will only operate when the parameters described within the ODD are satisfied.” *(SAE, 2021)*

AD can just be realized in certain operational design domains and their restrictions

Operational Design Domain:

*"…**operating conditions under which a given driving automation system or feature thereof is specifically designed to function**, including, but not limited to, environmental, geographical, and time-of-day restrictions, and/or the requisite presence or absence of certain traffic or roadway characteristics."*

SAE (2018)

Exemplary ODD restrictions:

Weather:
System can just function with clear sight and at daylight conditions

Geography:
System can just function in clear defined geofenced area (e.g. specific city area)

Driving Speed:
System can just function in driving speeds up to 60 km/h

Roadway:
System can just function on streets with structurally separated driving lanes where no pedestrians have access

Each company can define its own ODDs under which its self-driving systems can operate. There is no standardization in place and therefore ODDs can vary from system to system.

Level 5 is a mission impossible – driving anytime, anyplace would require Human-Level Intelligence

LEVELS OF DRIVING AUTOMATION

AREA OF POSSIBLE APPLICATION

Description	**Driver support features – You are driving** (e.g. warning, emergency braking, lane keeping assist, traffic jam pilot)		**Automated driving features – You are not driving** (e.g. L3 highway pilot, L4 driverlass taxi)			
SAE Levels	L1	L2	L3	L4	L4	L5
	Private Ownership			**Access Economy**		
Driver Responsibility	**Driver has to drive** when system requests to			**Driver is able to drive**	**Driverless** (no driver intervention in the car needed or possible)	
Environmental Restrictions	Restricted to **specific ODD*** (not anywhere)					On-road (anywhere), **driver-manageable**
Level of Intelligence	**Task-based AI** (Narrow)**					**General AI****
Liability	**Driver is liable**		**OEM is liable**			

WAYMO LEVEL 4 DRIVERLESS TAXI

(HD mapped city area + remote operation center)

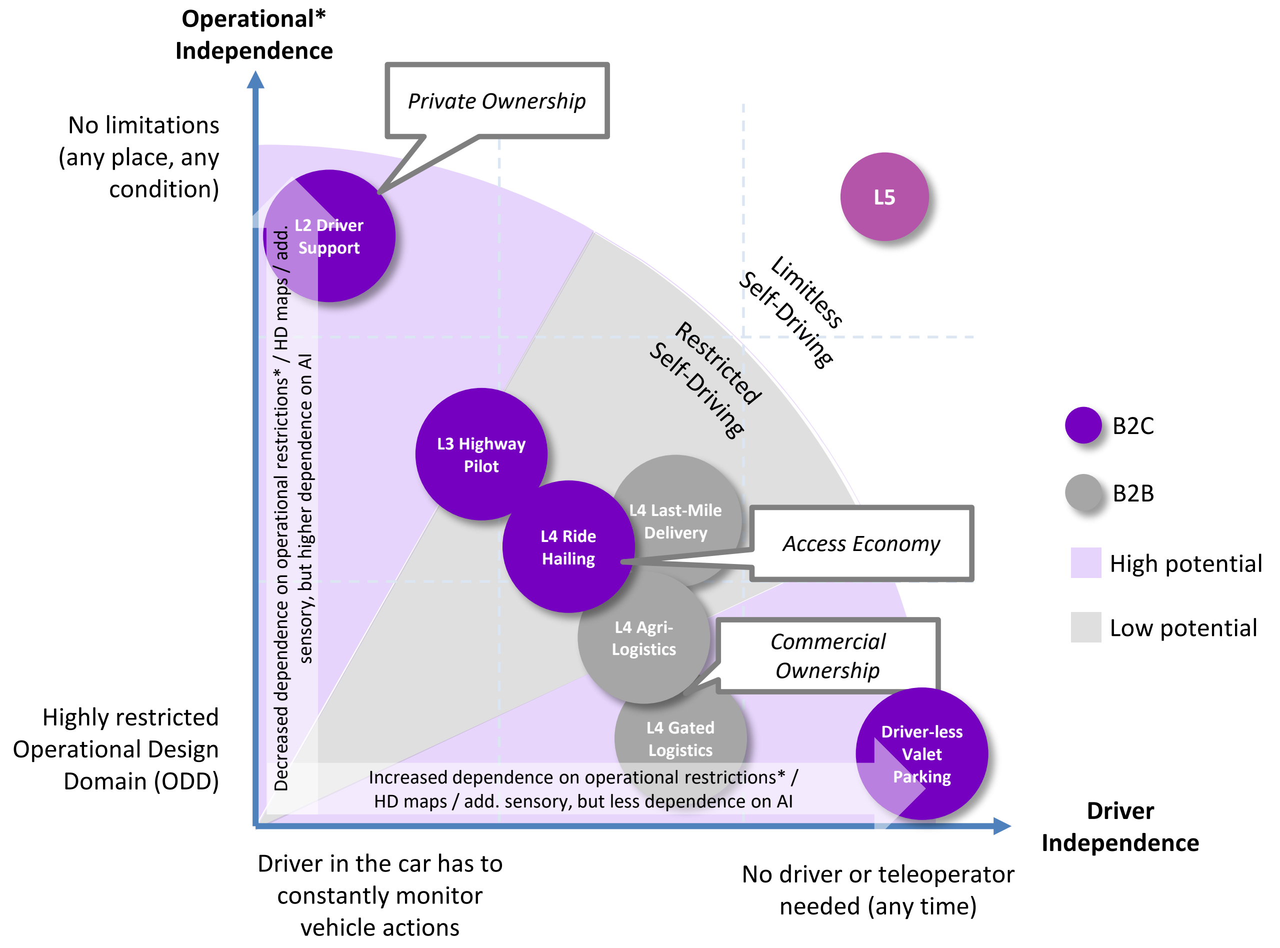
MERCEDES-BENZ LEVEL 3 DRIVE PILOT

(specific highways, only up to 60 km/h)

* Operational Design Domain: Operating conditions under which a given driving automation system or feature is specifically designed to function including geographic & environmental restrictions + traffic & roadway characteristics

** Artificial Intelligence

Although no level 5 – AD has several high potential business applications



* *Geographic & environmental restrictions + traffic & roadway characteristics*

GOOD NEWS FOR ESTABLISHED OEMS

Ownership model still intact

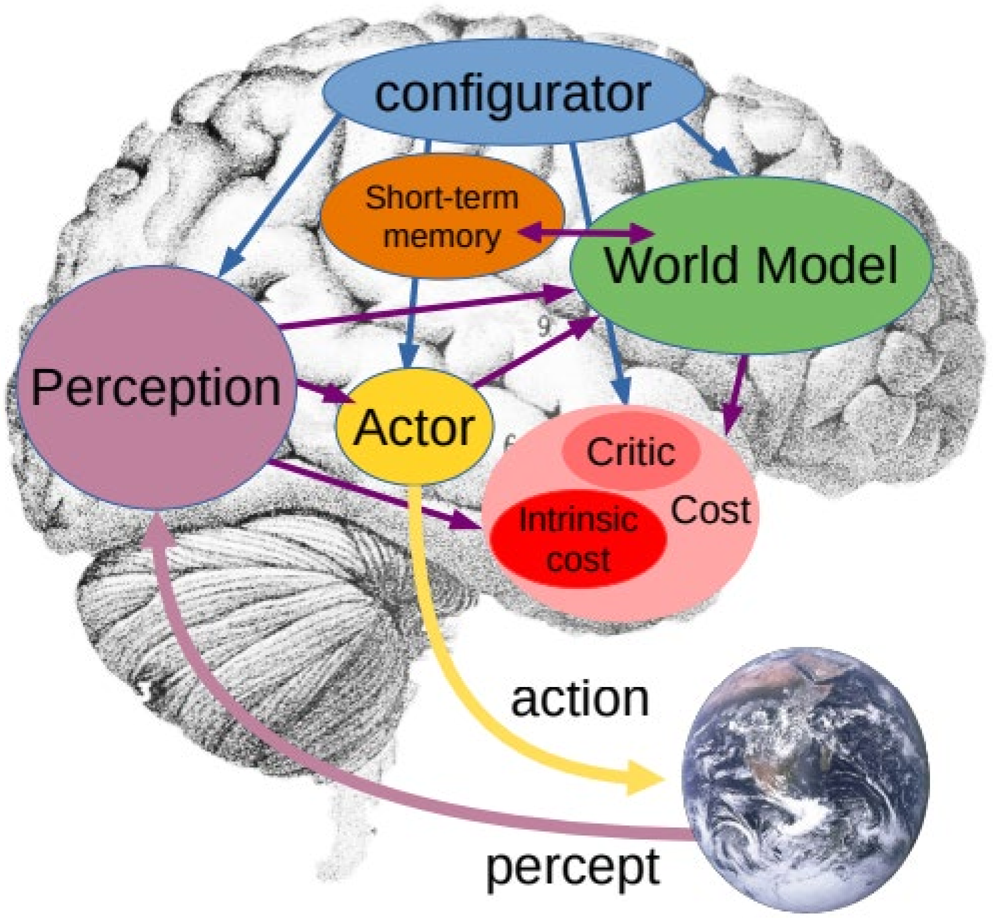
Monetization of incremental AD features

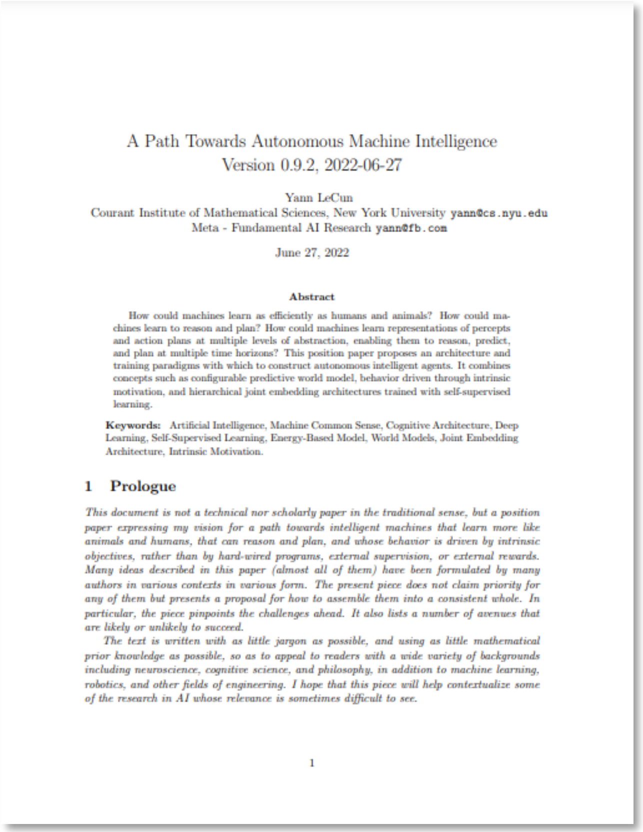
Large fleets to harvest a diverse set of big data

Thank You!

Backup

Latest cognitive models use multi-module architecture





A Path Towards Autonomous Machine Intelligence
Version 0.9.2, 2022-06-27

Yann LeCun
Courant Institute of Mathematical Sciences, New York University yann@cs.nyu.edu
Meta - Fundamental AI Research yann@fb.com

June 27, 2022

Abstract

How could machines learn as efficiently as humans and animals? How could machines learn to reason and plan? How could machines learn representations of percepts and action plans at multiple levels of abstraction, enabling them to reason, predict, and plan at multiple time horizons? This position paper proposes an architecture and training paradigms with which to construct autonomous intelligent agents. It combines concepts such as configurable predictive world model, behavior driven through intrinsic motivation, and hierarchical joint embedding architectures trained with self-supervised learning.

Keywords: Artificial Intelligence, Machine Common Sense, Cognitive Architecture, Deep Learning, Self-Supervised Learning, Energy-Based Model, World Models, Joint Embedding Architecture, Intrinsic Motivation.

1 Prologue

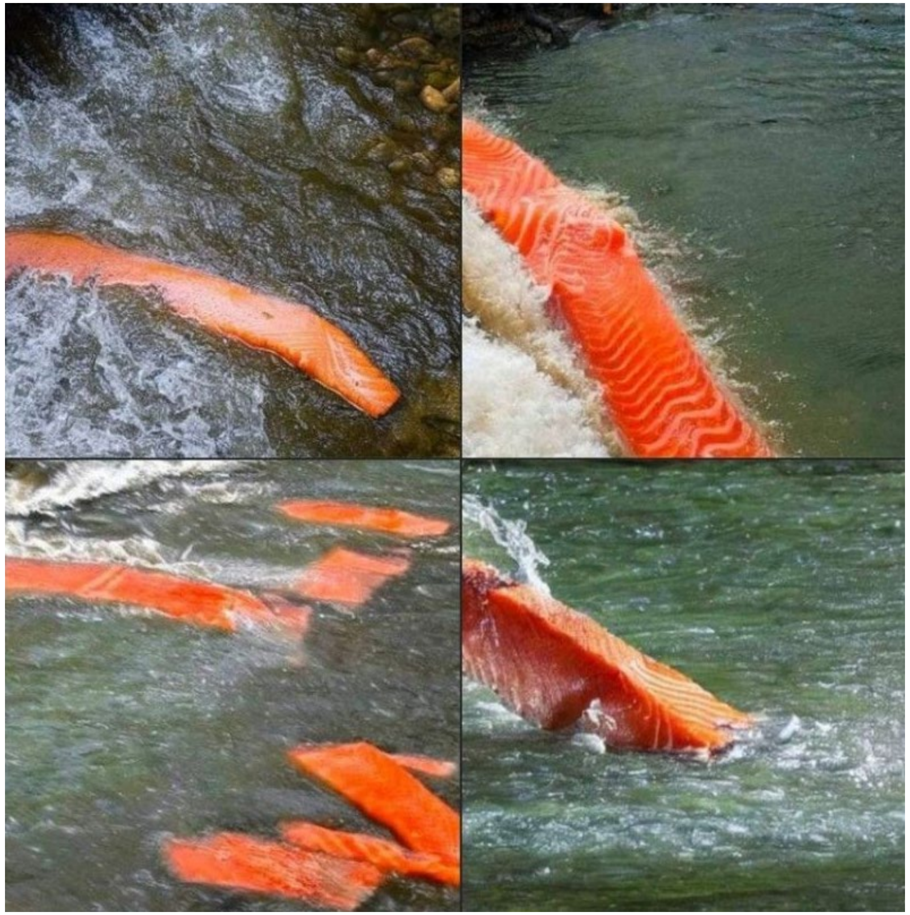
This document is not a technical nor scholarly paper in the traditional sense, but a position paper expressing my vision for a path towards intelligent machines that learn more like animals and humans, that can reason and plan, and whose behavior is driven by intrinsic objectives, rather than by hard-wired programs, external supervision, or external rewards. Many ideas described in this paper (almost all of them) have been formulated by many authors in various contexts in various form. The present piece does not claim priority for any of them but presents a proposal for how to assemble them into a consistent whole. In particular, the piece pinpoints the challenges ahead. It also lists a number of avenues that are likely or unlikely to succeed.

The text is written with as little jargon as possible, and using as little mathematical prior knowledge as possible, so as to appeal to readers with a wide variety of backgrounds including neuroscience, cognitive science, and philosophy, in addition to machine learning, robotics, and other fields of engineering. I hope that this piece will help contextualize some of the research in AI whose relevance is sometimes difficult to see.

1

Yann LeCun (2022) - [Link]

Barrier of meaning – text to speech



Tobias Gutmann (2022)

Text to speech: “A salmon swimming down the river“

A text-to-image model is a machine learning model that takes a **natural language description as input and produces an image that matches that description**.

The **development of such models began in the mid-2010s** as a result of advances in deep neural networks, and is advancing quite fast.

However, of course, there are still many shortcomings - as this funny image illustrates.

Training of AI models is based on brute force computing requiring large volume of data

LARGE/SPECIFIC DATASETS

Huge amount of real-world data with specific properties is required for training a neural net image classifier:

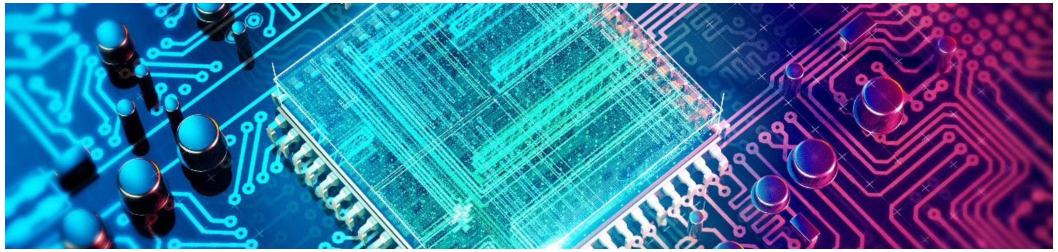
- Training data typically **10x bigger** than number of parameters of a neural net
- Data is **split between training and testing** to reduce overfitting
- **High variety** of example within **various classes** is needed

Example:
ImageNet (dataset for computer vision): **14 million** labelled images, **21,841 subcategories**, **27 high level categories**

Gonfalonierie (2019), Kelly (2016), Shrivastava (2017), Seif (2018), Hu Xu, et al. (2018)

+

BRUTE FORCE COMPUTING



Major AI achievements are often based on sheer computing power:
The amount of computational power used to train the largest AI models **doubles every 3.4 months**

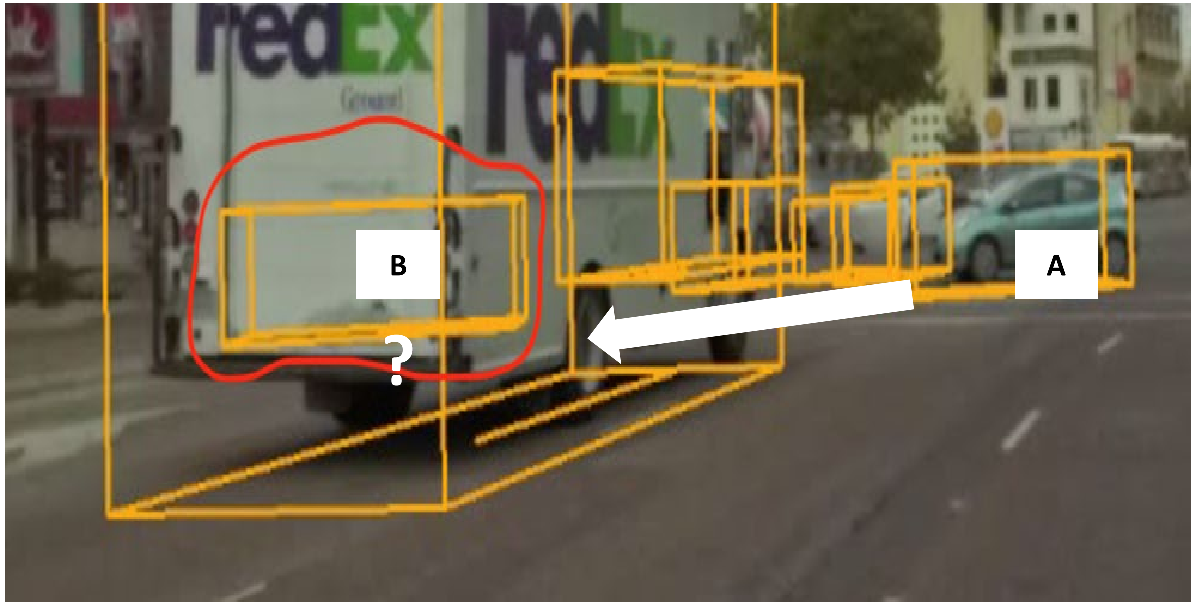
Example:
Facebook requires **huge processing power worth $4,1 million in hardware** (32 CPUs + 256 P100 GPUs) for one-hour training of a neural net image classifier based on ImageNet

OpenAI (2018), You et.al (2018)

AI models are hungry for data and computational power and try to simulate quality through quantity

AI systems struggle with object permanence, so tracking an object over time

TO FOLLOW THE COURSE OF THE CAR YOU NEED OBJECT PERMANENCE & OBJECT TRACKING



Without spatial and physical understanding, the **car stops to exist as soon as it is occluded** by the truck. **Humans think and memorize in scenes** (episodes) and thus are familiar with the case of temporarily occluded objects.

Four obstacles to common sense reasoning:

1. Occluded object vision
2. Object permanence
3. Imagination of the unseen
4. Episodic memory

Object permanence becomes important in dynamic environments. It cannot be studied in still images. Once videos are concerned, the problem rapidly gains importance. **Autonomous driving is "the literally killer problem with object permanence"**.

(Ken Ryu, Medium, 2016)