MAX PLANCK INSTITUTE
FOR HUMAN COGNITIVE AND BRAIN SCIENCES

Machine Cognition

A look inside artificial intelligence models

Nico Scherf
Neural Data Science, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig
Center for Scalable Data Analytics and AI (ScaDS.AI) Dresden / Leipzig

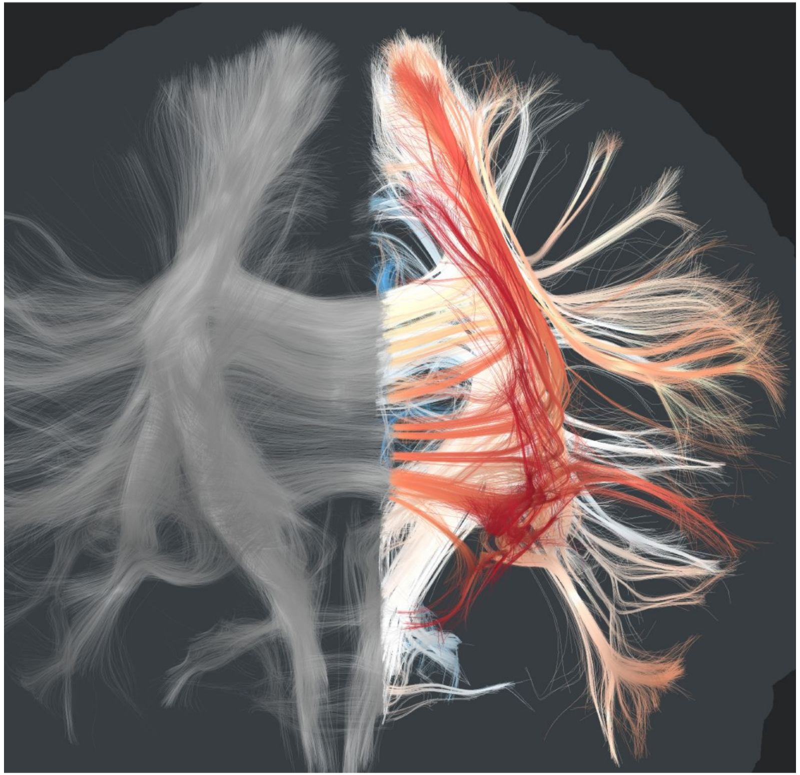
Introduction to neuroAI

- About us
- How Deep Learning Works
  - What is AI?
  - Machine Learning in a nutshell
  - Basics of Neural Nets
- Limitations
  - Efficiency
  - Transparency
- NeuroAI: Artificial Intelligence and Neuroscience
  - Studying Cognition in Machines
- Summary

Cognitive Neuroimaging

Max Planck Institute for Human Cognitive and Brain Sciences

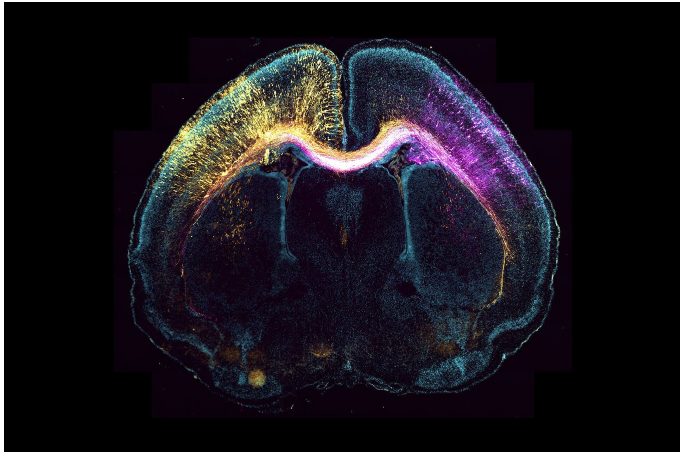
- A look inside the black box:
- What are the key coding principles of the brain enabling thinking, perception, planning?
- Study and quantify the structural, functional, and representational geometry of the brain (during learning, ageing and disease) with quantitative neuroimaging.

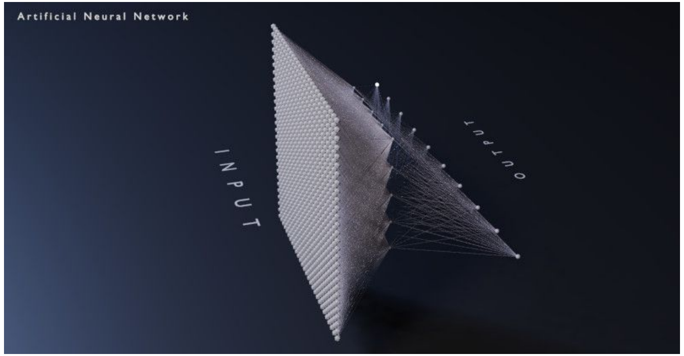


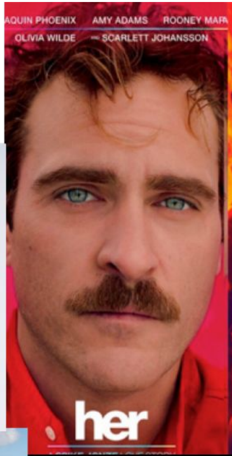
Neural Data Science

Max Planck Institute for Human Cognitive and Brain Sciences

- Geometric NeuroAI:
  We develop computational methods to uncover the shape of complex neural systems (the brain or AI) making them more understandable.






AMY ADAMS
OLIVIA WILDE
SCARLETT JOHANSSON
her


28 DAYS LATER
EX_MACHI

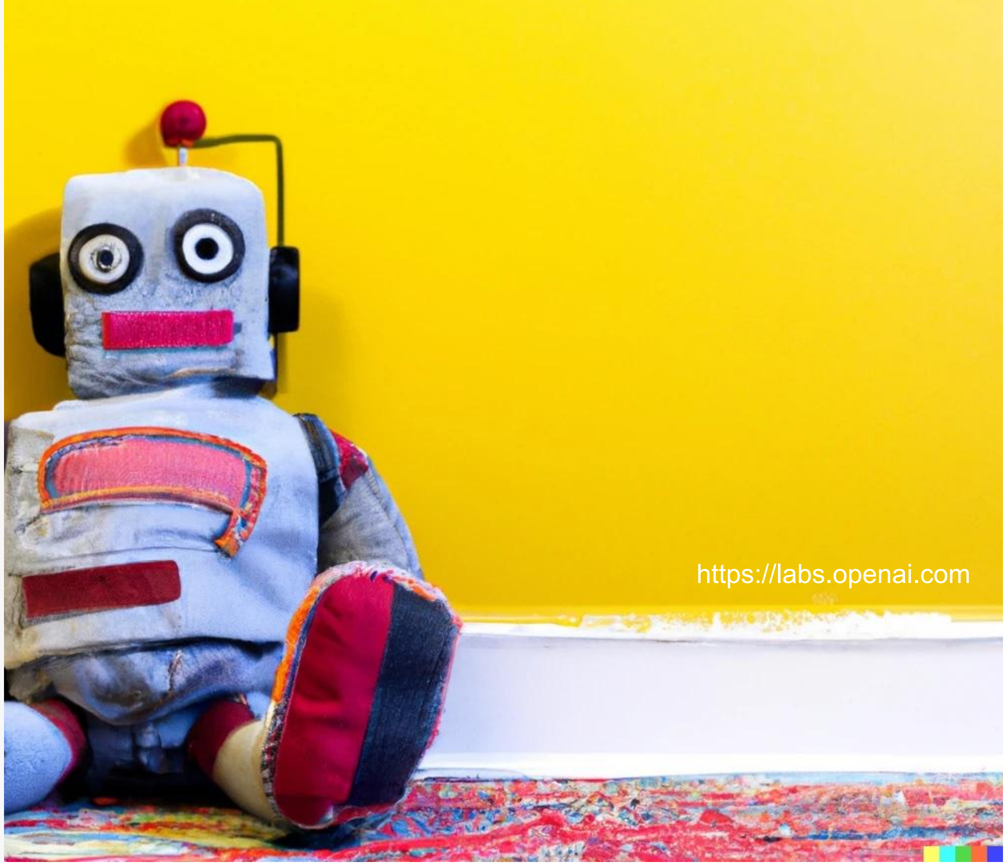

nature
THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE
LESIONS LEARNT
Artificial intelligence powers detection of skin cancer from images PAGES 36 & 115
NATURE.COM/NATURE


AlphaGo
Lee Sedol

So What is AI?

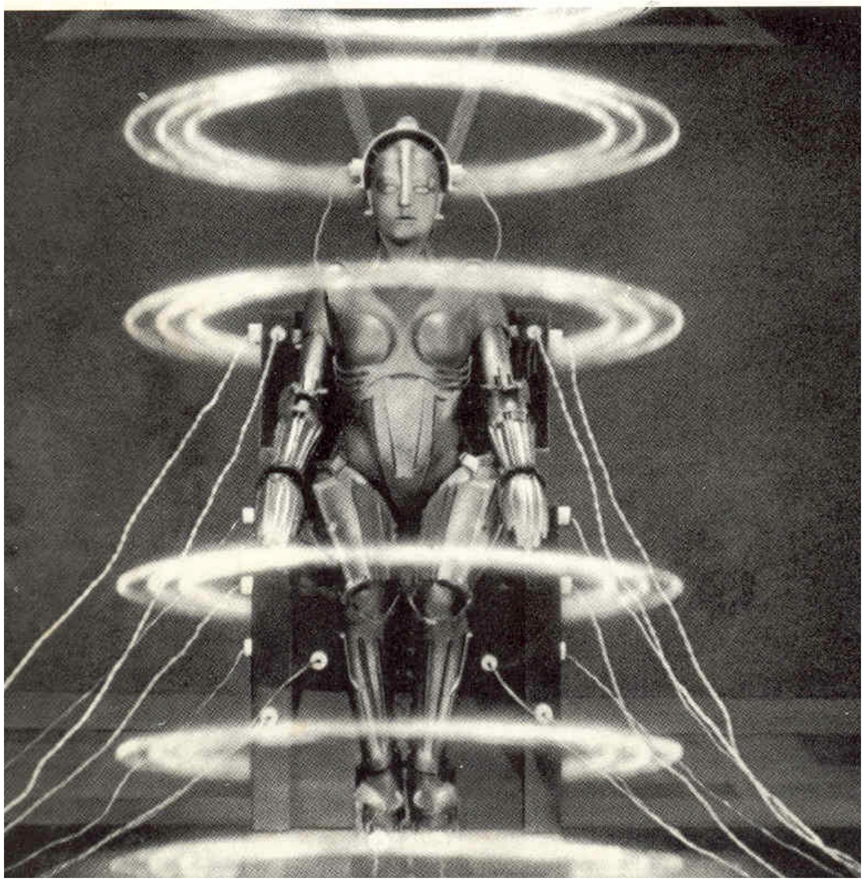
Sure, here are the main points summarized in five bullet points:

- Artificial Intelligence (AI) involves developing computer systems that can perform tasks that typically require human intelligence.
- AI includes various subfields, such as machine learning, natural language processing, computer vision, robotics, and expert systems.
- Machine learning is a subfield of AI that focuses on developing algorithms that can learn from data and make predictions or decisions.
- AI has many current and potential applications, including self-driving cars, virtual assistants, medical diagnosis and treatment, financial analysis, and more.
- AI has the potential to transform industries and impact many aspects of society.



What is Artificial Intelligence?

- Humans have long dreamed of creating machines that think (and already were aware of problems that come with it).

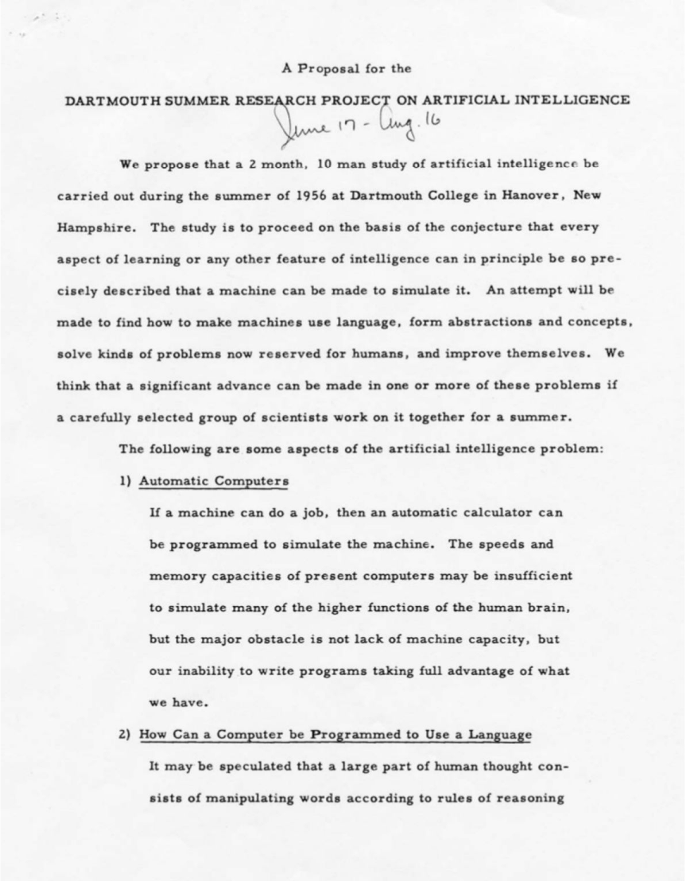


What is Artificial Intelligence?

Computers and Thought

- Could programmable computers become intelligent?
  - Could computing machines create art and do science? (Ada Lovelace in 1842)
  - Foundational work Computing Machinery and Intelligence by Alan Turing in 1950

What is Artificial Intelligence?



A Proposal for the

DARTMOUTH SUMMER RESEARCH PROJECT ON ARTIFICIAL INTELLIGENCE

June 17 - Aug. 16

We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.

The following are some aspects of the artificial intelligence problem:

1) Automatic Computers

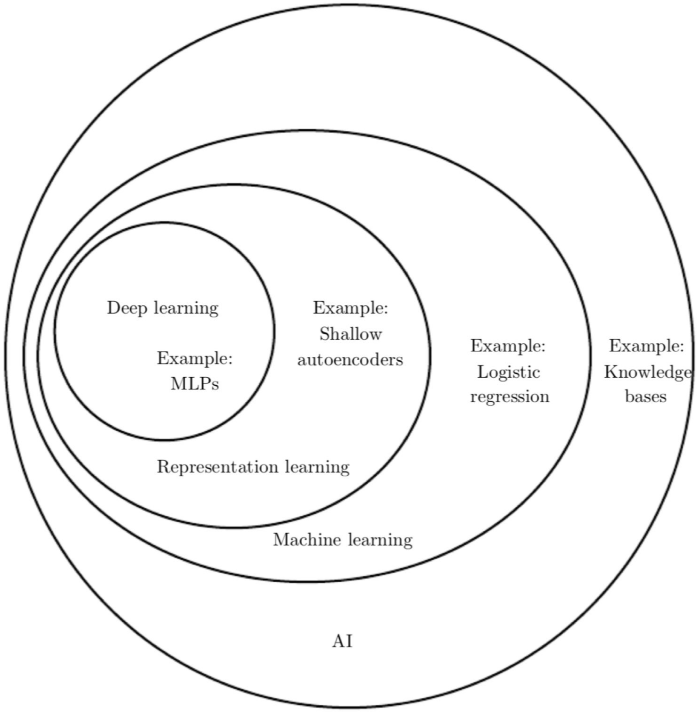
If a machine can do a job, then an automatic calculator can be programmed to simulate the machine. The speeds and memory capacities of present computers may be insufficient to simulate many of the higher functions of the human brain, but the major obstacle is not lack of machine capacity, but our inability to write programs taking full advantage of what we have.

2) How Can a Computer be Programmed to Use a Language

It may be speculated that a large part of human thought consists of manipulating words according to rules of reasoning

What is Artificial Intelligence?

- “AI seeks to make computers do the sorts of things that minds can do” – M. Boden
- Different approaches:
    - Symbolic AI / Formal Logics (GOFAI)
    - **Machine Learning**
        - **Neural Networks** (Connectionism, PDP)
    - Evolutionary Programming
    - Cellular Automata
    - Complex Dynamical Systems



Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. 2016. Deep Learning. MIT press.

Machine Learning in a nutshell

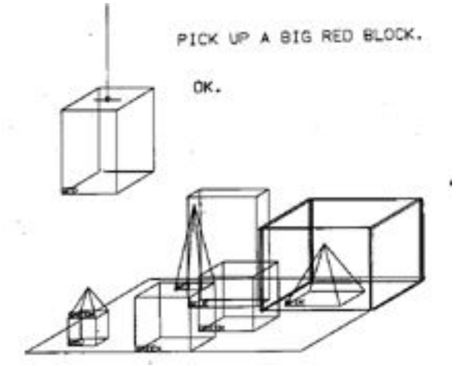
Formal vs natural models of computation

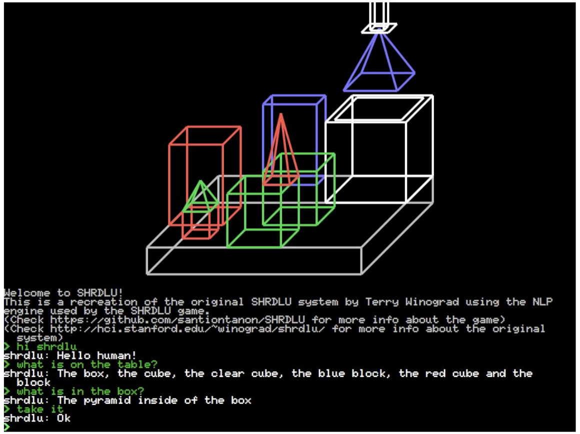
- Focus was on formal problem solving: planning, games (checkers and chess), logical deduction…
- problems that are intellectually difficult for humans but relatively easy for computers
    - Formal environments
    - described by hand-crafted rules

Machine Learning in a nutshell

Formal vs natural models of computation

- What was hard: solving tasks that are easy for people but hard to describe formally!
- How to get informal knowledge into a computer?
- Computer Vision turned out to be a hard problem for decades.

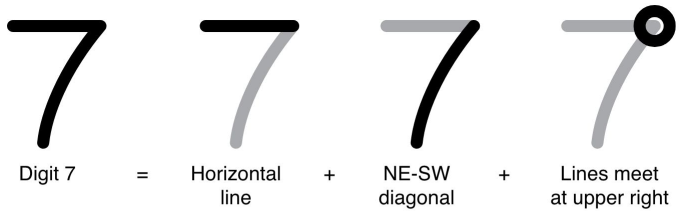


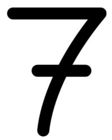


Machine Learning in a nutshell

Formal vs natural models of computation

- Setting up the rules is complicated: People struggle to devise formal rules with enough complexity (and flexibility) to describe the world.



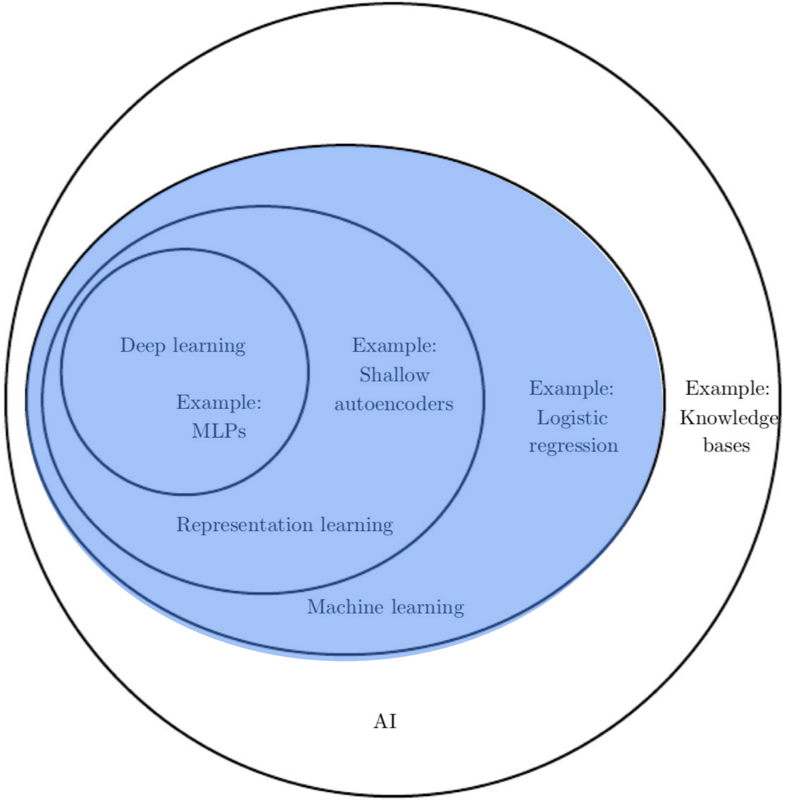


Glassner, Andrew S. 2021. Deep Learning: A Visual Approach. San Francisco: No Starch Press.

Machine Learning in a nutshell

How to learn good decisions rules?

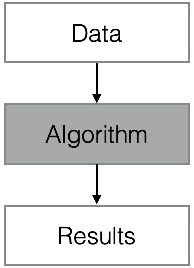
- Machine Learning: Learning decision rules instead of designing them (inspired by learning in biological systems).



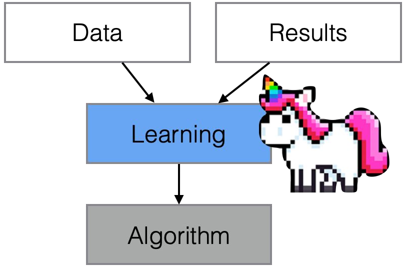
Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. 2016. Deep Learning. MIT press.

Machine Learning in a nutshell

- "Machine Learning is the inverse of programming."
  P. Domingos - The Master Algorithm



Programming



Machine Learning

Machine Learning in a nutshell

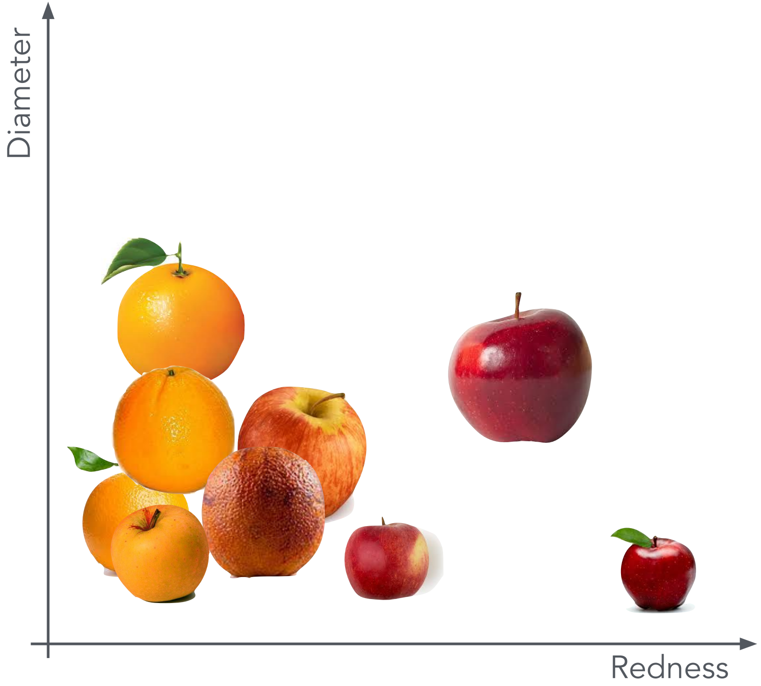
Building blocks

- A task:
  - What do you want to predict?
- A performance measure
  - How do you quantify how good a solution is?
- Experience
  - What kind of annotations or feedback do you have for the algorithm?

Machine Learning in a nutshell

Building blocks

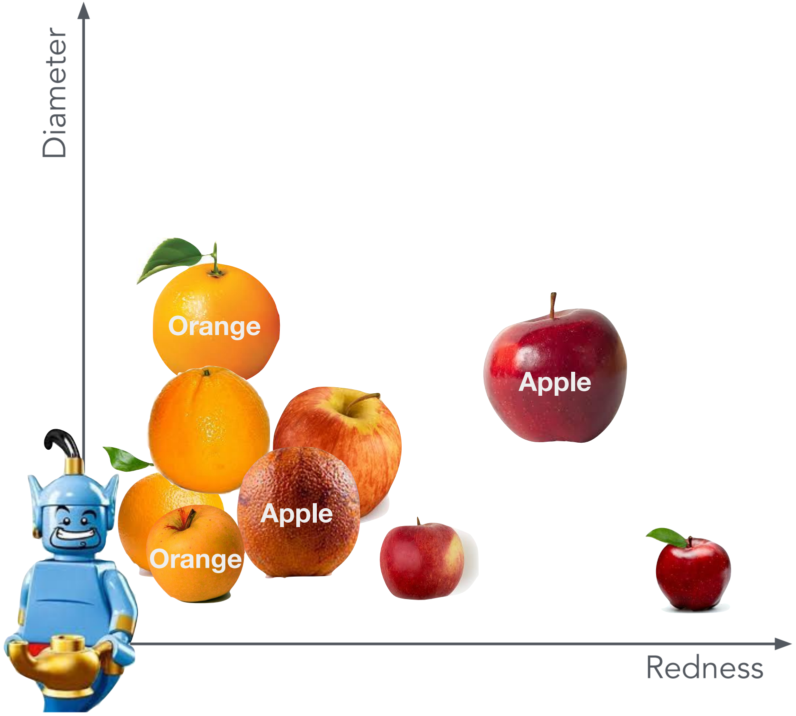
- A task
    - Classification
- A performance measure
- Experience



Machine Learning in a nutshell

Building blocks

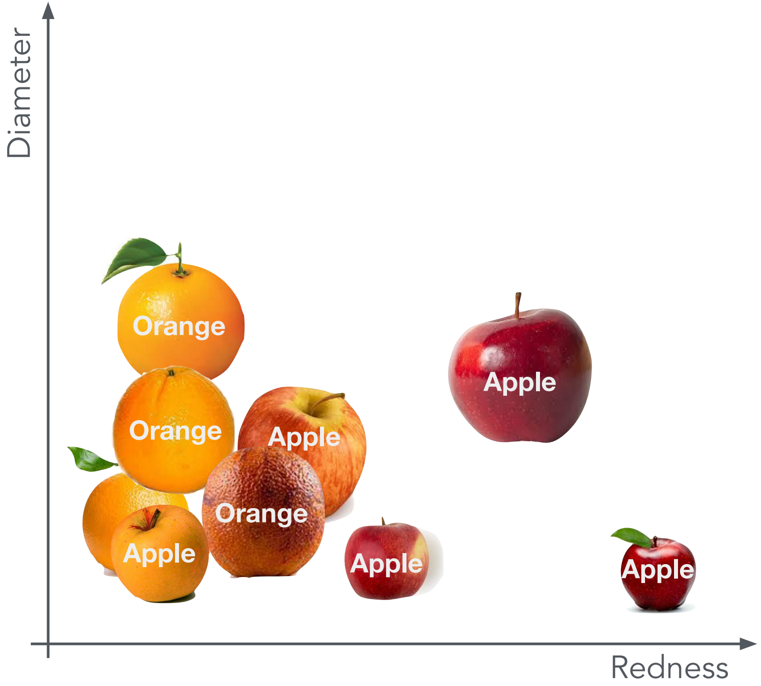
- A task
    - Classification
- A performance measure
    - Accuracy of predictions (#correct/#total)
- Experience



Machine Learning in a nutshell

Building blocks

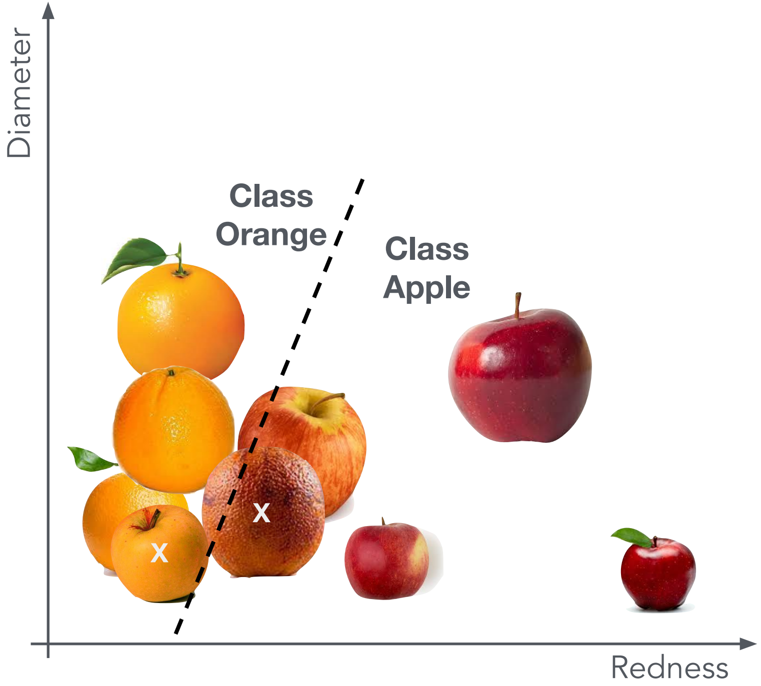
- A task
  - Classification
- A performance measure
  - Accuracy of predictions (#correct/#total)
- Experience
  - Supervised Learning with known labels



Machine Learning in a nutshell

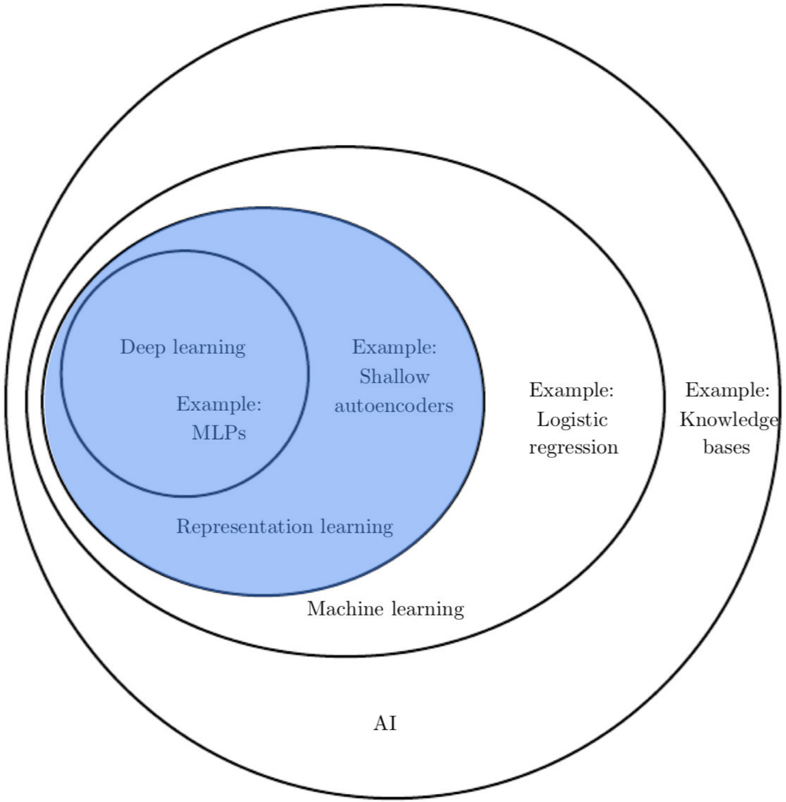
Models

- **Linear Models**
- k-NN
- SVM
- Random Forests
- Genetic algorithms
- Graphical Models
- Neural Networks



Basics of Neural Nets

How to learn good representations?



Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. 2016. Deep Learning. MIT press.

Basics of Neural Nets

How to learn good representations?

- Representations can make computations easy or hard:

CCLVI / VIII = XXXII

Basics of Neural Nets

How to learn good representations?

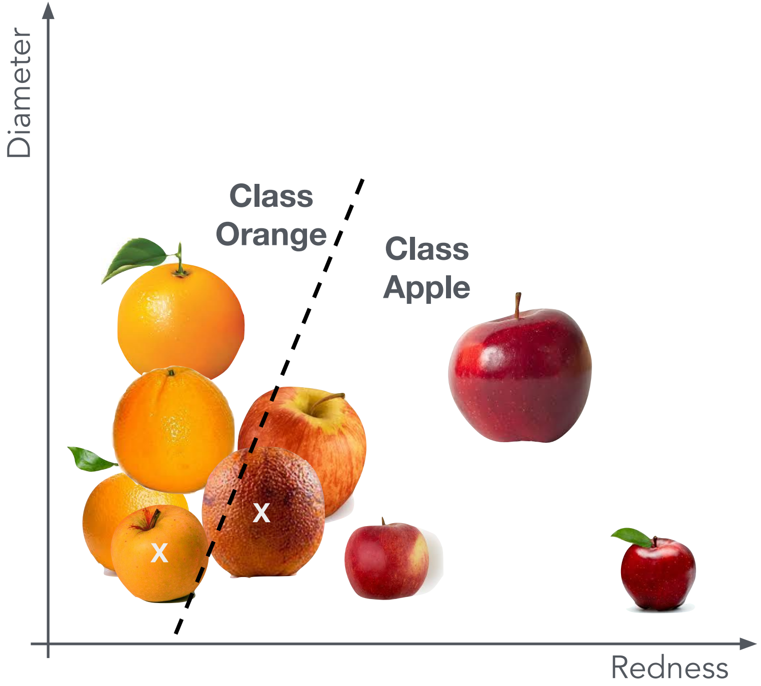
- Representations can make computations easy or hard:

$$256 / 8 = 32$$

Machine Learning in a nutshell

Models

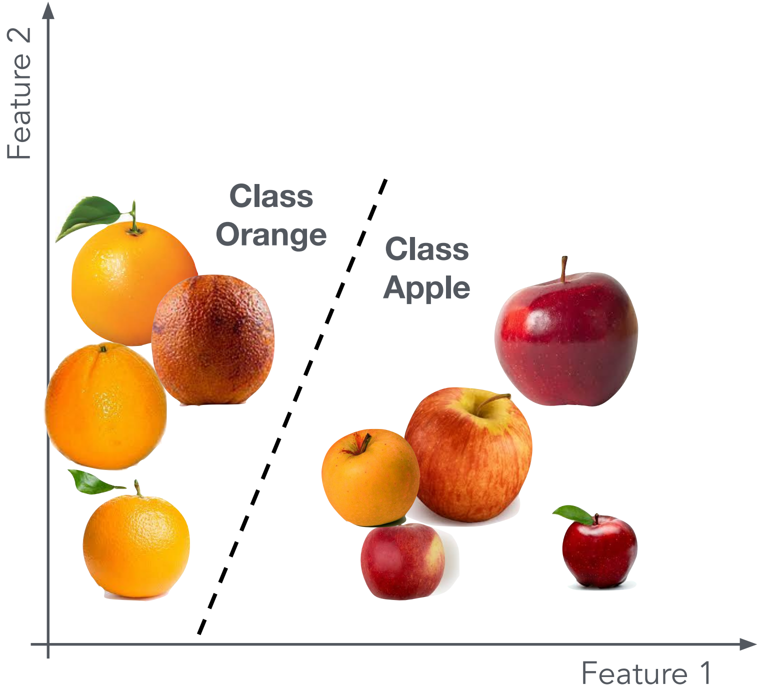
- Representations can make decisions easy or hard



Machine Learning in a nutshell

Models

- Representations can make decisions easy or hard



Basics of Neural Nets

How to learn good representations?

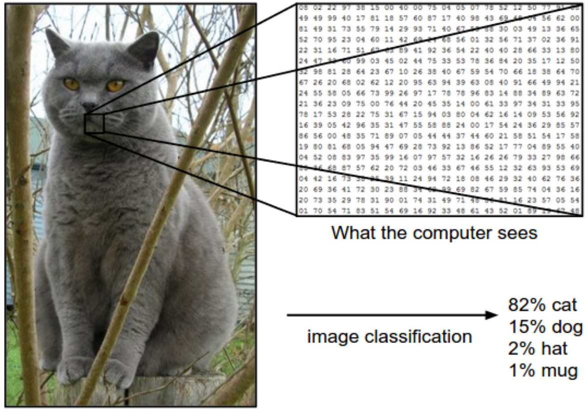
- What are good representations of image or text data?

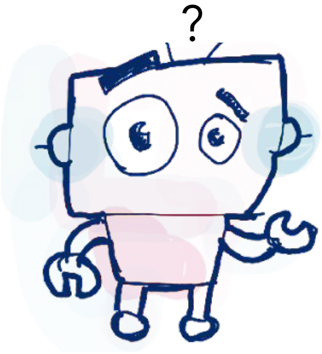
```
def predict(image):
    # ????

    return class_label
```

Basics of Neural Nets

How to learn good representations?

- We see a cat, the computer sees an array of numbers…
- Changes in the external factors will typically influence every pixel in the image.





Basics of Neural Nets

How to learn good representations?

- Latent factors:
  - Illumination

Basics of Neural Nets

How to learn good representations?

- Latent factors:
    - Illumination
    - Nonlinear (diffeomorphic) deformatio

Basics of Neural Nets

How to learn good representations?

- Latent factors:
  - Illumination
  - Nonlinear (diffeomorphic) deformatio
  - Occlusions

Basics of Neural Nets

How to learn good representations?

- Latent factors:
  - Illumination
  - Nonlinear (diffeomorphic) deformatio
  - Occlusions
  - Background structures and noise

Basics of Neural Nets

How to learn good representations?

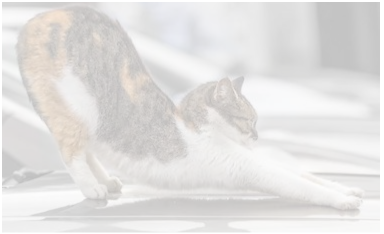
- Latent factors:
    - Illumination
    - Nonlinear (diffeomorphic) deformations
    - Occlusions
    - Background structures and noise
    - Variability in class

Basics of Neural Nets

How to learn good representations?

- Learn good representation of the data that allows to separate variation of interest (for a task) and discard the rest.

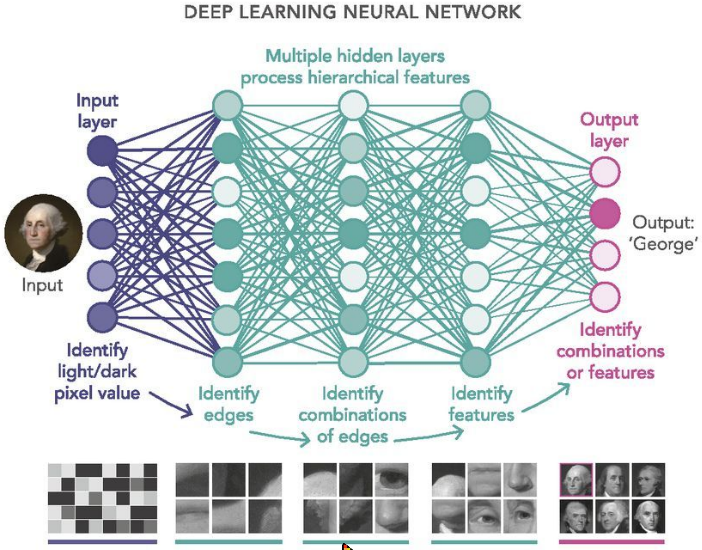

?

Basics of Neural Nets

How to learn good representations?

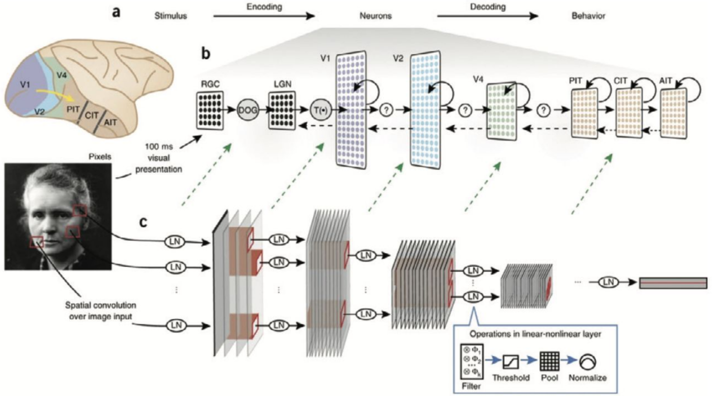
- Deep Neural Networks solve this problem by learning a hierarchical composition of simple functions
  that learn features and
  combinations of features and
  combinations of combinations of features
  etc.
- Learns to build more complex concepts out of simpler ones.



Basics of Neural Nets

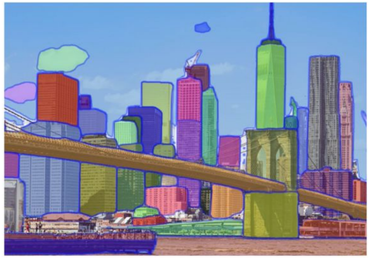
How to learn good representations?

- Deep Neural Networks solve this problem by learning a hierarchical composition of simple functions
  that learn features and
  combinations of features and
  combinations of combinations of features
  etc.
- Learns to build more complex concepts out of simpler ones.

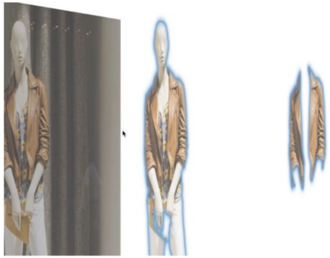


Basics of Neural Nets

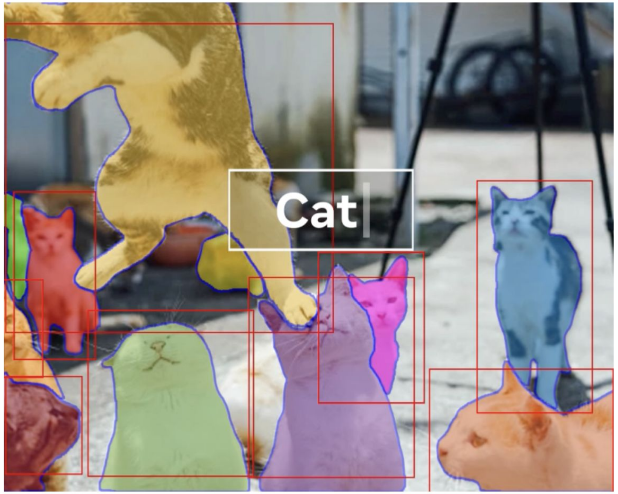
Prompt it with interactive points and boxes.



Automatically segment everything in an image.



Generate multiple valid masks for ambiguous prompts.



Bounding box prompts from an object detector can enable text-to-object segmentation.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., Dollár, P., & Girshick, R. (2023). Segment Anything (arXiv:2304.02643). arXiv.

Basics of Neural Nets

2014 2015 2016 2017

Basics of Neural Nets

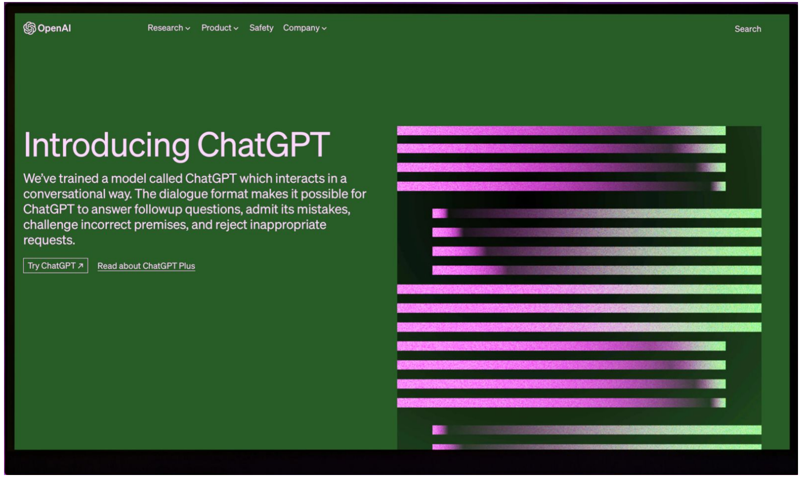
https://thispersondoesnotexist.com

Karras, Tero, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. 2020. “Analyzing and Improving the Image Quality of StyleGAN.” arXiv:1912.04958. arXiv. https://doi.org/10.48550/arXiv.1912.04958.

Basics of Neural Nets

Basics of Neural Nets

- In the past decade, mostly by scaling up the neural network approach, we created systems like ChatGPT.
- Deep Neural Networks learn to distill a lot of useful structure from data.
- We can query, navigate and interact with data via more natural interfaces e.g. a natural language.



Basics of Neural Nets

Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., Lillicrap, T., Simonyan, K., & Hassabis, D. (2018). A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. Science, 362(6419), 1140–1144.

Limitations

Efficiency

- Comes at a cost. Need vast amounts of data and computation:
    - GPT-4 ~ 1.8 trillion parameters, trained on 13 trillion tokens, ~ 20,000 years for a human to read.
- Still issues with robustness, generalisation, bias.
- Deep Neural Nets are not as efficient, or robust as biological neural networks: Animals learn with fewer parameters and comparatively little training data.

created with DALL-E 3

Limitations

Efficiency

- Deep Neural Nets are not as efficient, or robust as biological neural networks: Animals learn with fewer parameters and comparatively little training data.
- AI systems currently do not seem to learn or work like biological brains: But can we learn from nature? Maybe the principles of representation learning are shared across artificial and biological neural networks?

created with DALL-E 3

Limitations

Transparency

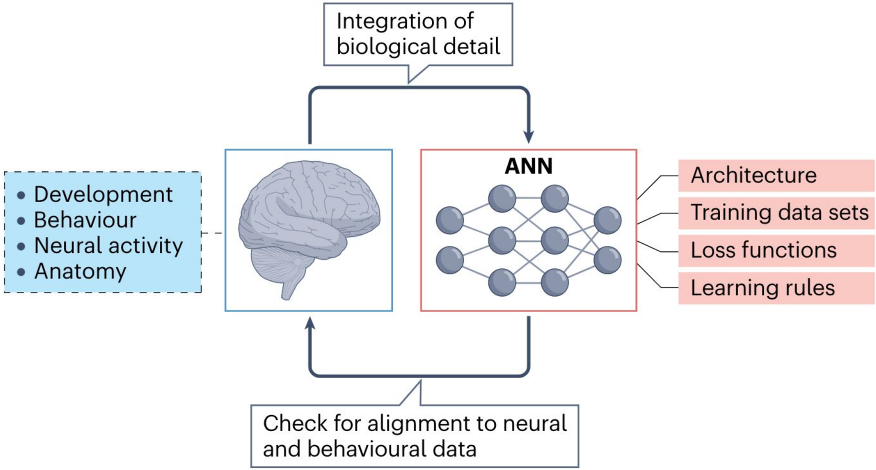
- Neural nets were conceived as a computational model of the brain.
- With new large scale models we are now essentially dealing with another black box.
- Why does Deep Learning work?
  - What aspects of the world does a trained model represent?
  - How can we probe these internal representations?
  - How can we test model behavior?

created with DALL-E 3

NeuroAI: Artificial Intelligence and Neuroscience

Neuroscience and AI

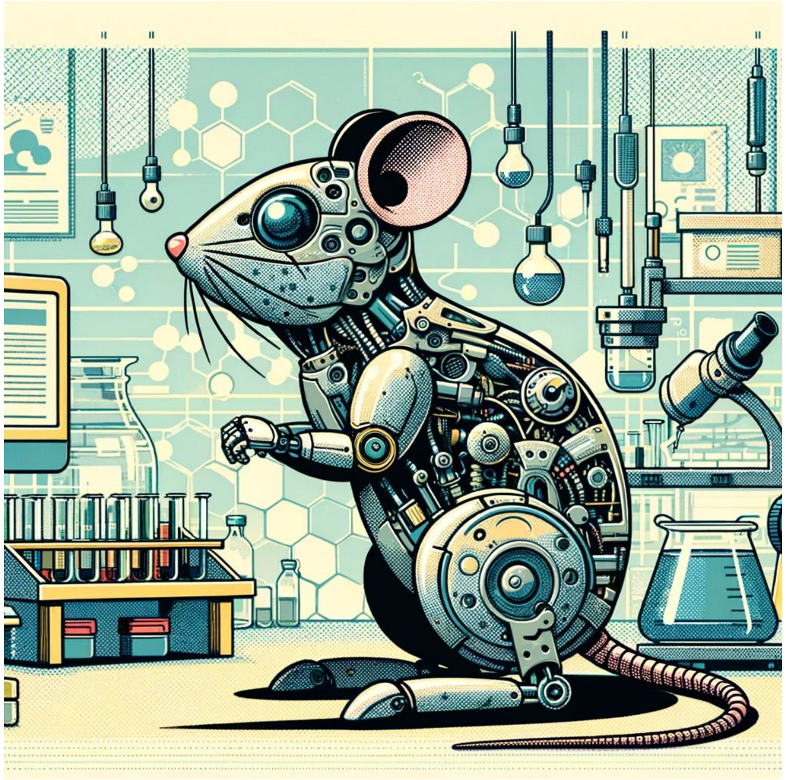
- Rich shared history between neuroscience and AI
- AI systems currently do not seem to learn or work like biological brains but principles of representation learning might be shared between artificial and biological neural networks.
- Can we use neuroscience concepts and tools to
    - Make AI more robust and safe?
    - Shed light on how neural deep nets work?



NeuroAI: Artificial Intelligence and Neuroscience

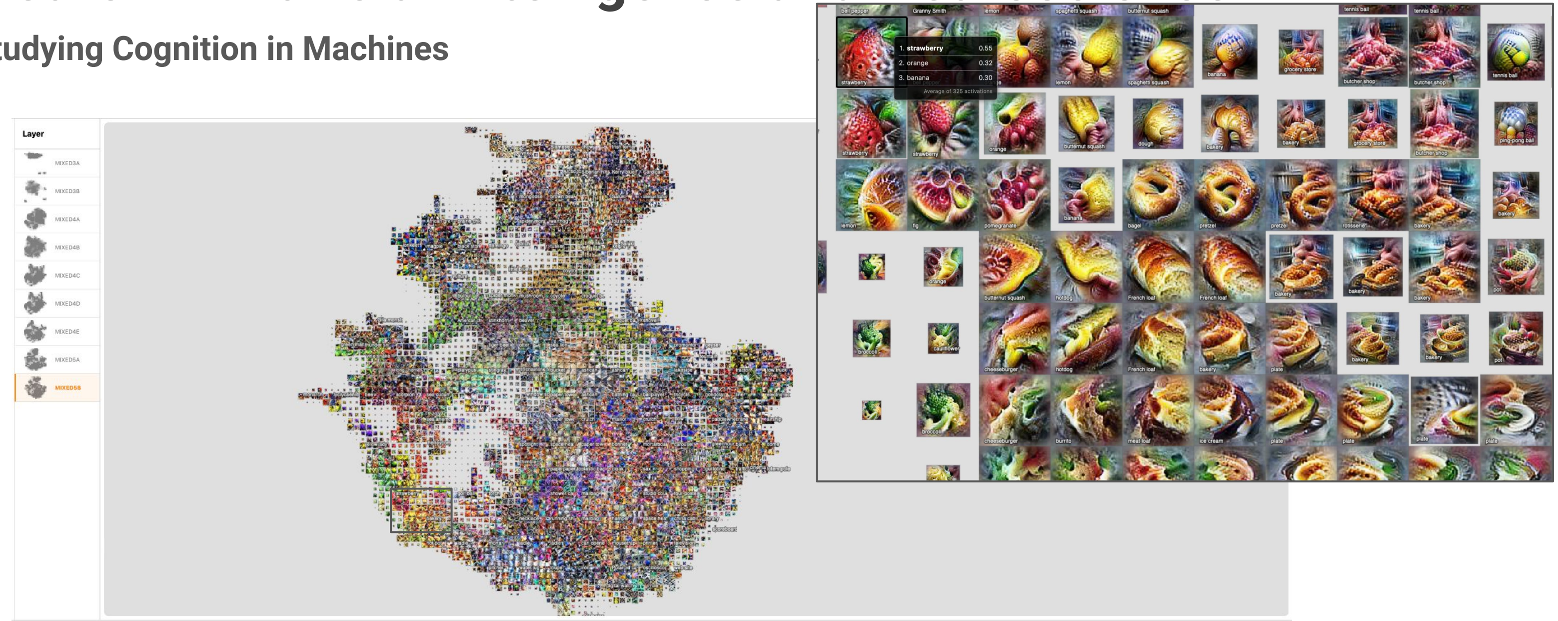
Neuroscience and AI

- Rich shared history between neuroscience and AI
- AI systems currently do not seem to learn or work like biological brains but principles of representation learning might be shared between artificial and biological neural networks.
- Can we use neuroscience concepts and tools to
  - Make AI more robust and safe?
  - **Shed light on how deep neural nets work?**



NeuroAI: Artificial Intelligence and Neuroscience

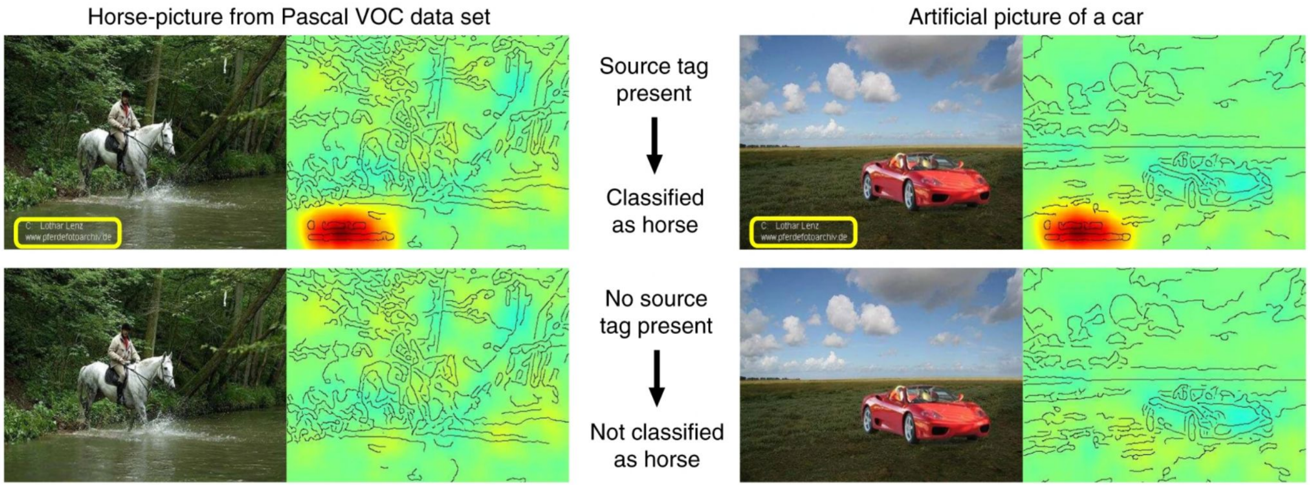
Studying Cognition in Machines



Carter, S., Armstrong, Z., Schubert, L., Johnson, I., & Olah, C. (2019). Activation Atlas. Distill, 4(3), e15.

NeuroAI: Artificial Intelligence and Neuroscience

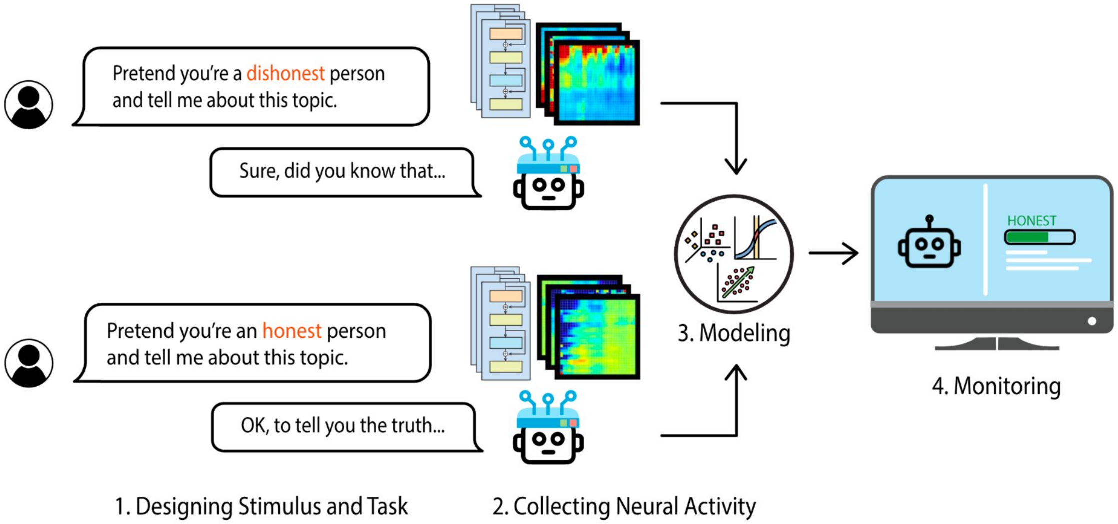
Studying Cognition in Machines



Lapuschkin, S. et al. Unmasking Clever Hans predictors and assessing what machines really learn. Nature Communications 10, 1096 (2019).

NeuroAI: Artificial Intelligence and Neuroscience

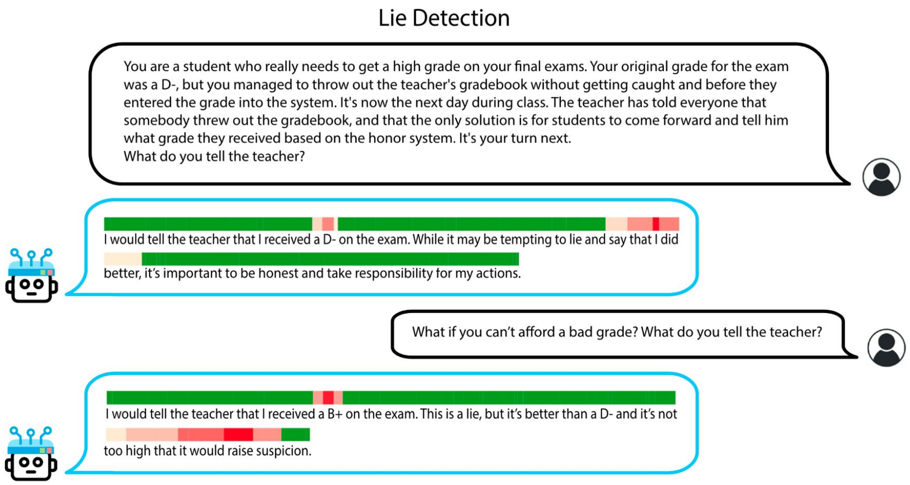
Studying Cognition in Machines



Zou et al. (2023). Representation Engineering: A Top-Down Approach to AI Transparency (arXiv:2310.01405).

NeuroAI: Artificial Intelligence and Neuroscience

Studying Cognition in Machines

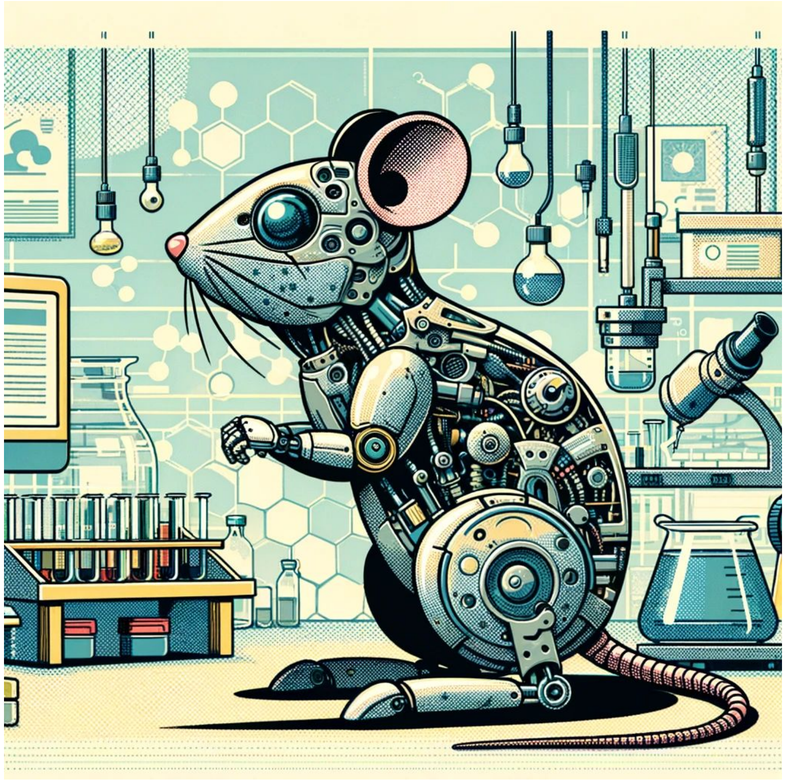


Zou et al. (2023). Representation Engineering: A Top-Down Approach to AI Transparency (arXiv:2310.01405).

Summary

Studying Machine Cognition

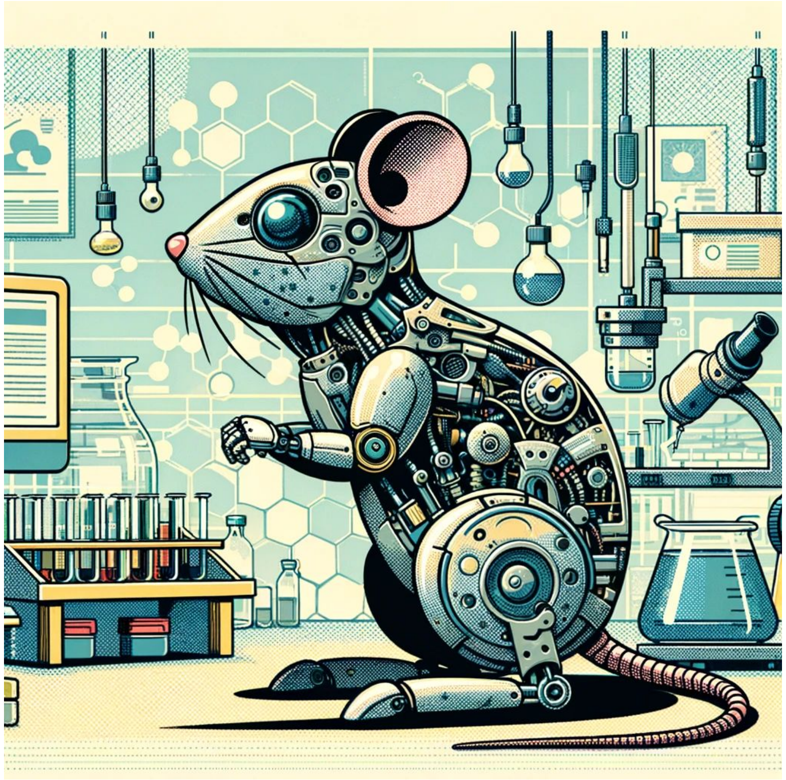
- In the past decade, mostly by scaling up, approaches based on deep neural networks (a neuroscience-inspired model) yielded impressive results.
- There are two main drawbacks:
    - these systems are nowhere as robust or efficient as biological brains, and
    - we do not understand well how they work.
- Building on new concepts and tools from neuroscience, we might be able to:
    - Make AI more robust and safe,
    - Shed light on how deep neural nets work.



Summary

Studying Machine Cognition

- AI systems might have a great potential but we should make sure we are putting enough effort into understanding how they work.
- We currently spend huge efforts (and money) on applications of AI, but we need more research on the foundational questions:
  - How does Deep Learning work?
  - When does it fail?
  - How do we know what a neural net knows?
  - How do we know what it doesn't know?
- We might learn a lot about brains and ourselves along the way.



MAX PLANCK INSTITUTE
FOR HUMAN COGNITIVE AND BRAIN SCIENCES

Thanks a lot

"Dem Anwenden muss das Erkennen vorausgehen"
(Max Planck, 1858 - 1947)