

Introduction to Deep Learning

Introduction and Overview



Slides of this lecture make use of / are based on:

- Andreas Geiger (University of Tübingen): Deep Learning (<https://uni-tuebingen.de/fakultaeten/mathematisch-naturwissenschaftliche-fakultaet/fachbereiche/informatik/lehrstuehle/autonomous-vision/lectures/deep-learning/>)
Here slide deck 1 (Introduction)

1.1 Welcome

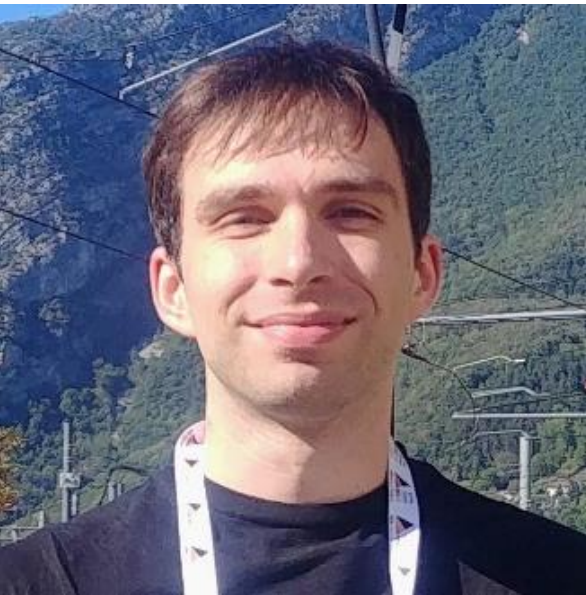




- Welcome
- Examples of Deep Learning
- Course Organization
- What is Deep Learning?
- History of Deep Learning



- Studied in Würzburg (Computer Science, Math, History)
- PhD in Computer Science (Data Mining) in Würzburg
- Postdoc at “Gesis – Leibniz Institute for the Social Sciences”, Cologne
- “Senior Postdoc” at RWTH 2017-2021
- Professor for Applied Machine Learning at Passau since April 2021



- Studied at RWTH Aachen
 - Electrical Engineering
 - Computer Science
- Research Assistant at University of Passau
 - Since October 2021

Machine Learning Methods

- Interpretable ML
- Rule
- Fairness
- Robustness
- Privacy

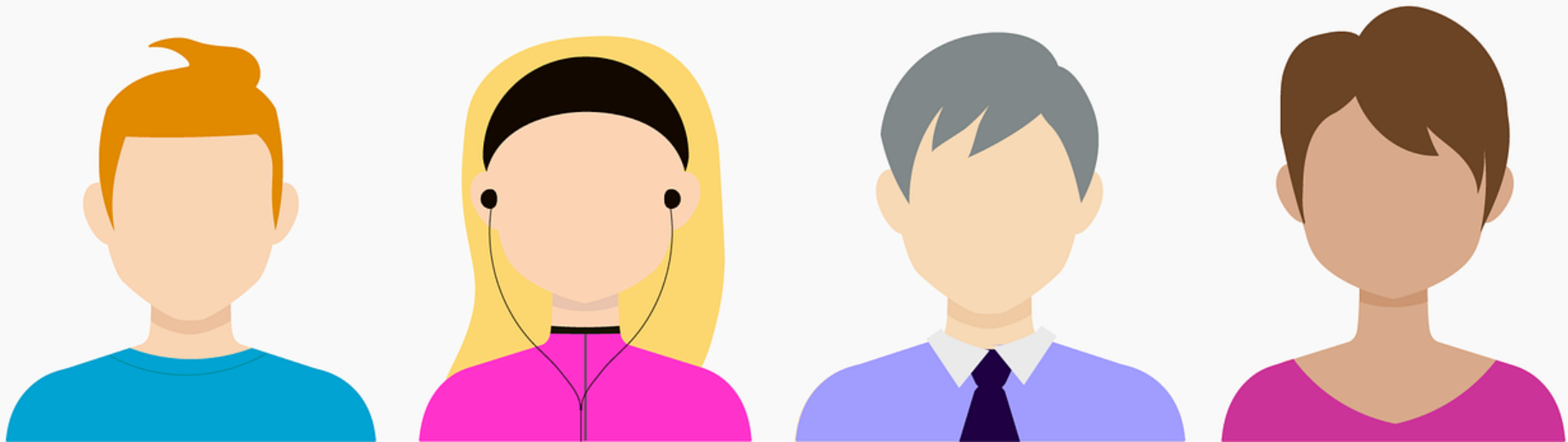
- Networks
- Text

Machine Learning Applications

- Use Machine Learning to analyze human behavior

- Wikipedia
- Twitter
- Reddit
- ...

Who are you?



1.2 Motivation: Examples of Deep Learning





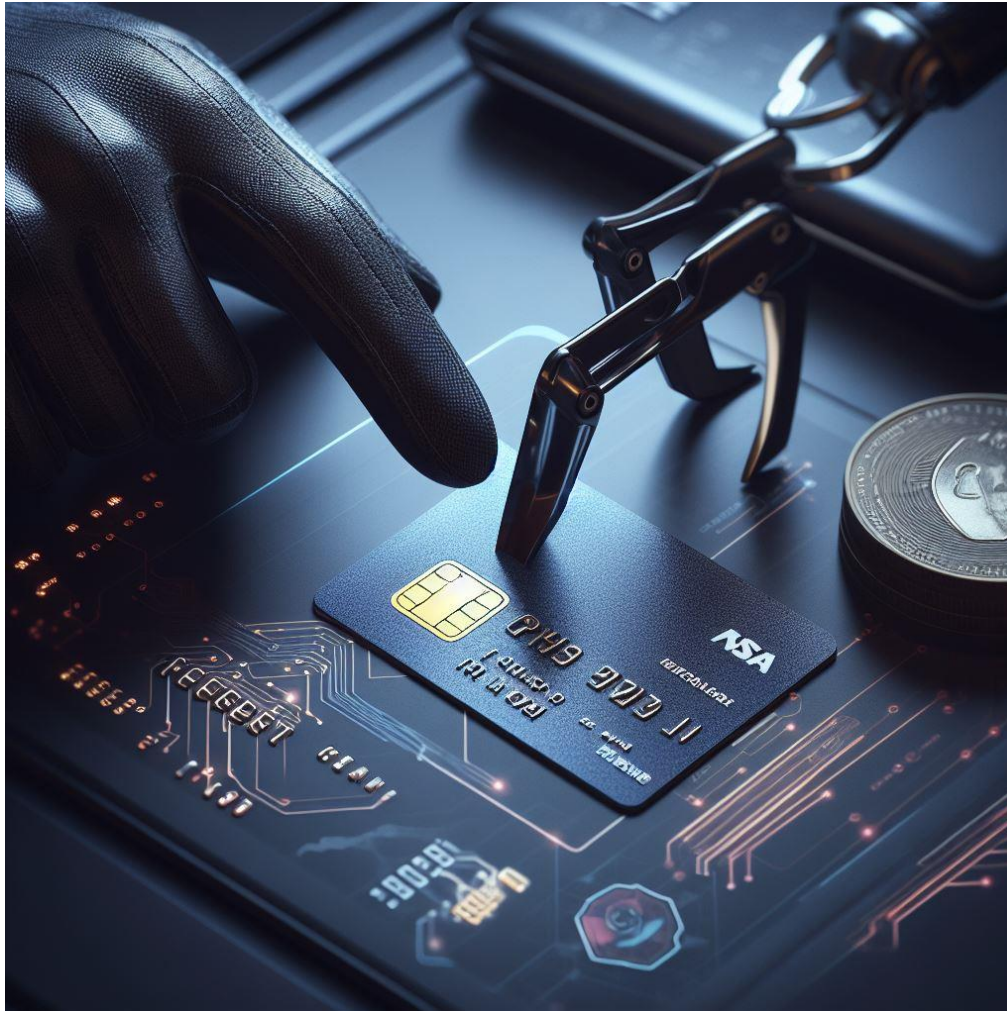




Image by Dall-E
3 via Bing Image
Generator




<https://github.com/lbeaucourt/Object-detection>


**Text übersetzen**
26 Sprachen

**Dateien übersetzen**
.docx und .pptx

Deutsch (erkannt) ▾

Herzliche Willkommen in Passau!
Ich begrüße Sie alle zu unserem Kurs zum
Thema "Deep Learning"!










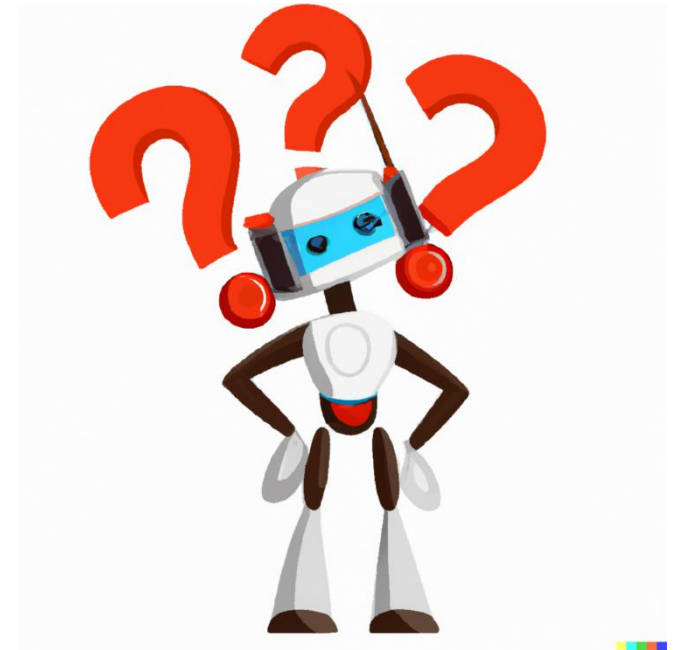
Englisch (US) ▾

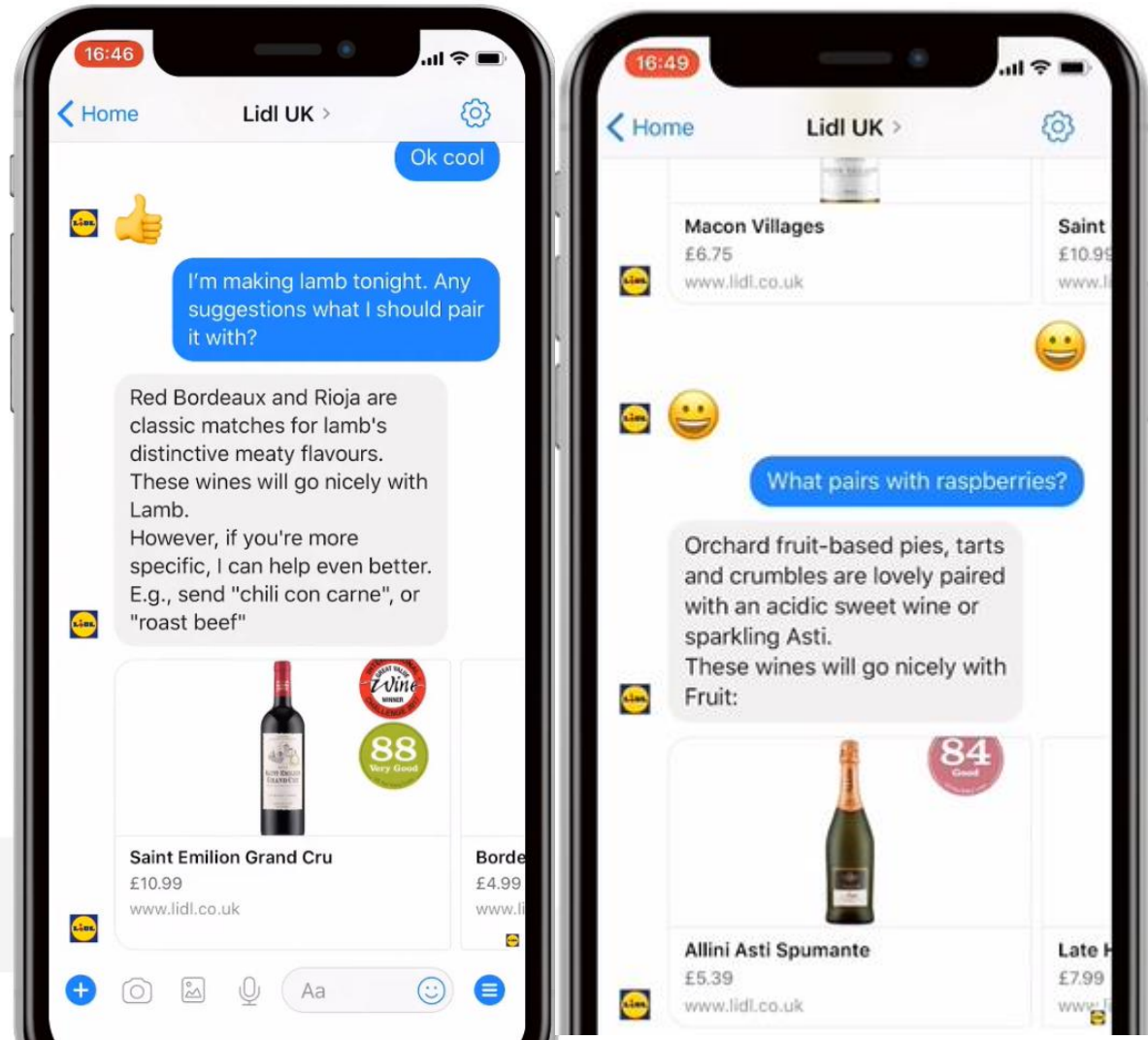
Glossar

Welcome to Passau!
I welcome you all to our course on "Deep
Learning"!



Screenshot from
translate.google.com





<https://www.chatbotguide.org/lidl-chatbot>

F Explain the backpropagation algorithm. Provide pseudo-code!



Sure! Backpropagation is a widely used optimization algorithm for training artificial neural networks. It involves computing the gradient of the loss function with respect to the weights of the neural network, and then using this gradient to update the weights in order to minimize the loss function during the training process. Here's a high-level overview of the backpropagation algorithm along with pseudo-code:



Input:

- Training dataset: X, Y
- Learning rate: α
- Number of iterations: `num_iterations`

Output:

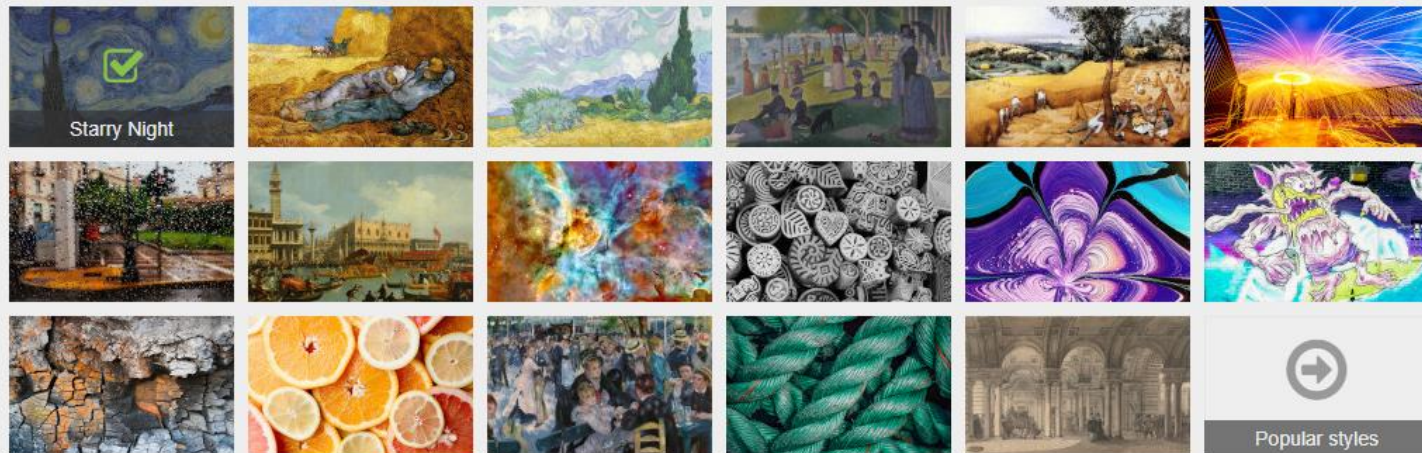
- Trained weights: W

Pseudo-code:

1. Initialize the weights of the neural network randomly or using some predefined method.
2. Loop for `num_iterations` times:

Screenshot from <https://chat.openai.com/>

Or select one of the default styles:



17





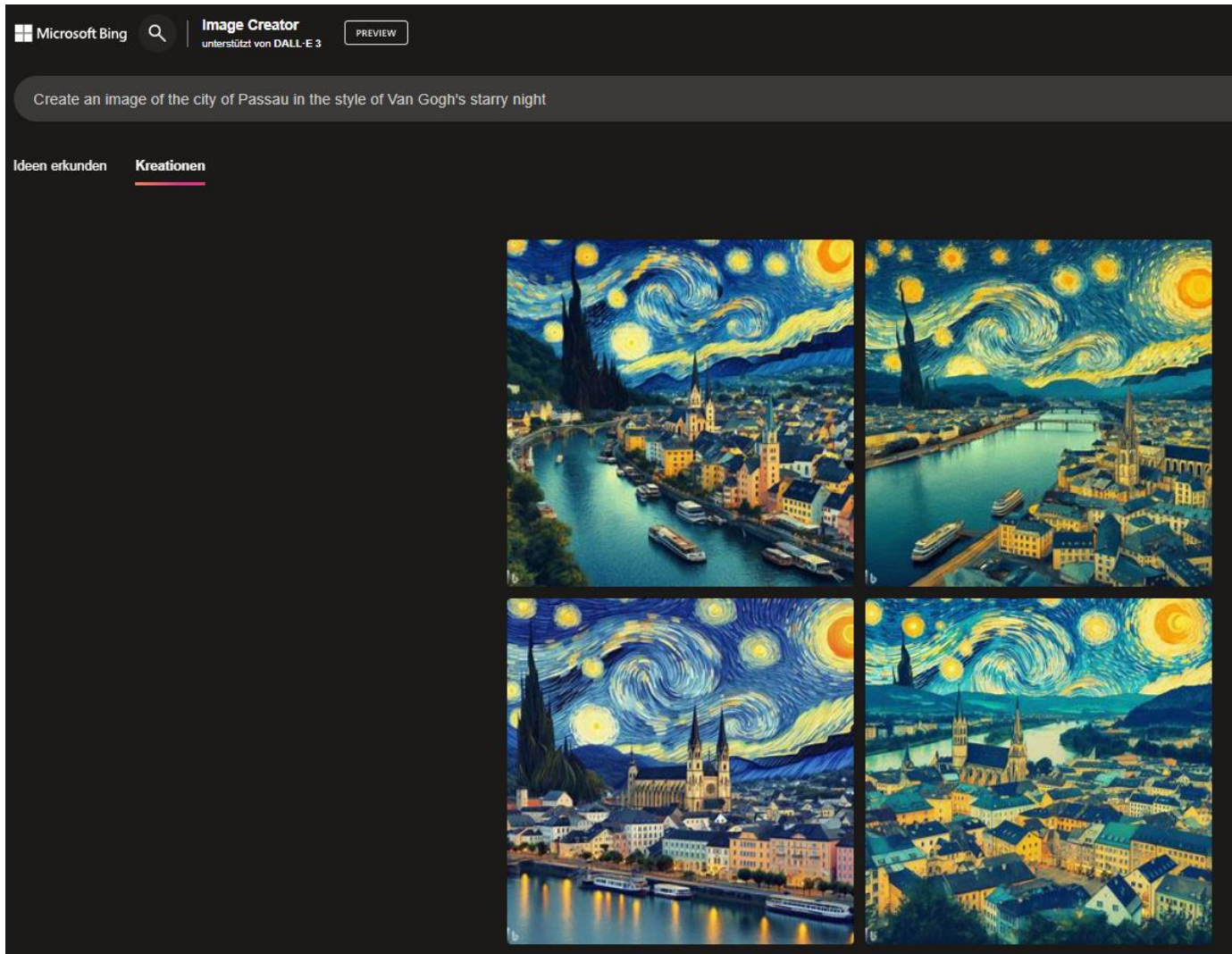


Image by Dall-E 3
via Bing Image
Generator



<https://github.com/junyanz/CycleGAN>

- Link prediction in social networks
- Credit card fraud detection
- Gender detection from text
- News aggregators (also with sentiment)
- AI that plays arcade games
- Image caption generation
- Grammar checker
- Inventing new algorithms(!)

1.3 Course Organization



- Understand the **fundamentals** of Deep Learning
- Know the basic concepts, terminology, algorithms
→ enable you to look up and understand new content quickly
- Enable you to apply Deep Learning in your own small projects
→ also do some implementations (in the exercises)
- Give a glimpse into current research
- ...but we are only scratching the surface

- Lectures: Tuesday 12.15h – 13.45h, (WIWI) HS 8
- Exercises: Two groups (identical content, attend one):
 - Thursday, 12.00h – 13.30h (WIWI) HS 8
 - Thursday, 14.00 – 15.30h, (WIWI) HS 8
- Language: English
- You should attend both, lectures and exercises to succeed in this course!
- Attending is not mandatory, but of course highly recommended
- Course slides will be put online in StudIP
- No intentional omissions, but completeness of the slide is not guaranteed
- „The spoken word counts“ (for exam content)
 - ➔ You should take notes

- There is lots of (excellent) material out there on the topic
- Advantage of this course: Interactivity!
- If something is unclear at the moment: ask immediately
- Additional opportunity to ask questions after each block
- Please use the forums to ask questions and help answering them
- (Almost) no question is a stupid question!

- Some things might be not optimal
- Some things might be wrong
- Please provide direct feedback!
- Official course evaluation later

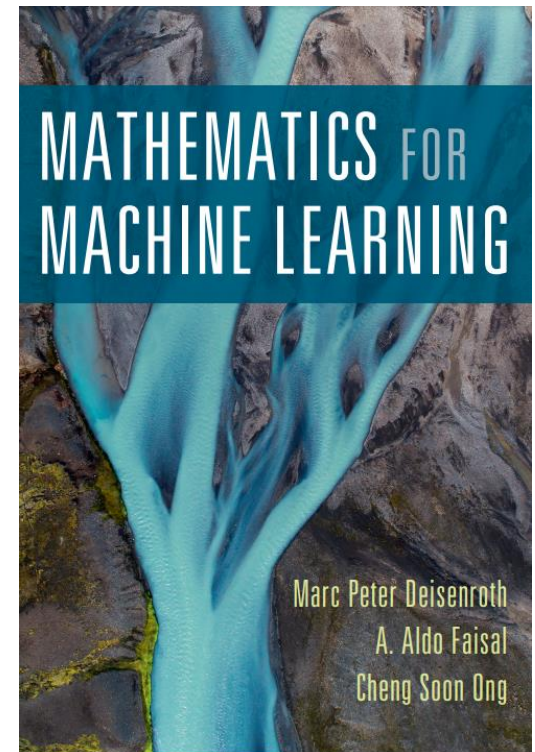
- Course under development, some part probably not in the shape as I would like to
- Experience:
Less questions, higher threshold for interactions with recordings
- Allows to speak „off-the-record“
- Please do not record privately

- (Very) tentative course program:

Session	Date	Topic	
1	18.10.2022	Introduction	
2	25.10.2022	History, Machine Learning	
3	8.11.2022	Machine Learning 2	
4	15.11.2022	Single-Layer Perceptrons	
5	22.11.2022	Single-Layer Perceptrons	
6	29.11.2022	Feed Forward	
7	6.12.2022	Feed Forward	
8	13.12.2022	Feed Forward	
9	20.12.2022	Convolutional Neural Networks	
11	10.1.2023	Recurrent Neural Networks	
12	17.01.2023	Foundation Models	
13	24.01.2023	Transformer	
14	31.01.2023	Autoencoder & GANs	
15	07.02.2023	Explainability	

- Exams will be written exams of 90 minutes
- No pre-requisites to participate (e.g. exercise scores)
- No bonus
- Do not forget to register in time!
- Will cover everything covered in the course:
 - Lectures and exercise
 - Might contain some coding
 - Including first (!) and last (!) lecture
- There will be a second exam before the beginning of the next semester
- Dates will be announced as soon as possible

- Linear Algebra, including:
 - Vectors, matrices,
 - Vector and matrix operations multiplications,
 - matrices as functions
- Statistics
 - Data types
 - (Conditional, joint) Probability/Likelihood,
 - Distributions
- Calculus
 - Minima, Maxima
 - Derivations, derivation rules
 - Partial Derivations & Gradients
- Recommendation:
<https://mml-book.github.io/book/mml-book.pdf>



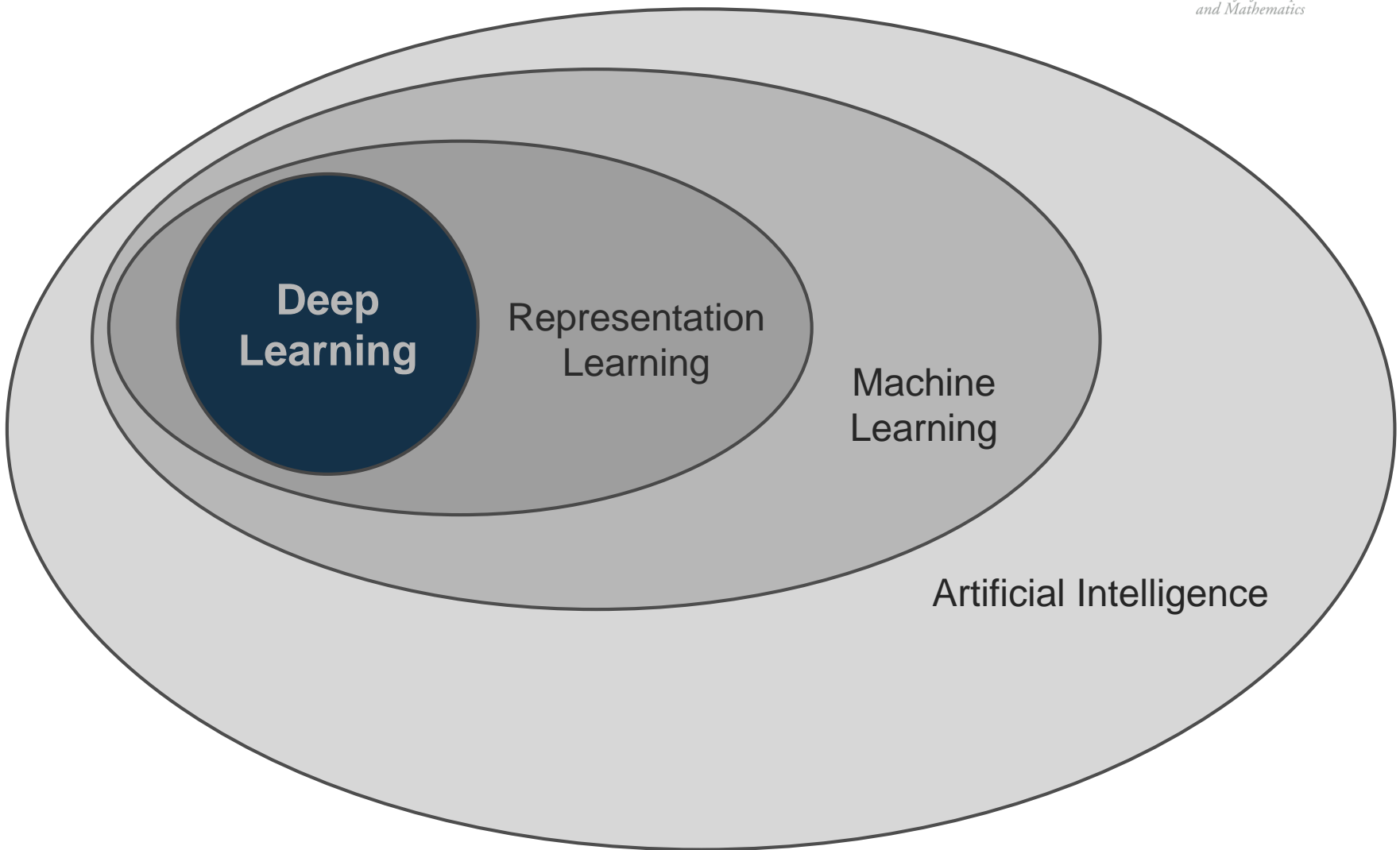
- Many (many, many) courses out there
 - Lecture: Deep Learning (Prof. Andreas Geiger, University of Tübingen)
<https://www.youtube.com/playlist?list=PL05umP7R6ij3NTWldtMbfvX7Z-4WEXRqD>
- ➔ Most closely to our course, material used from there

- Goodfellow, Bengio, Courville: Deep Learning
<https://www.deeplearningbook.org/>
- Aggarwal, Charu: Neural Networks and Deep Learning:
<http://www.charuaggarwal.net/neural.htm>
- Zhang, Lipton, Li, Smola: Dive into Deep Learning:
<http://d2l.ai/>

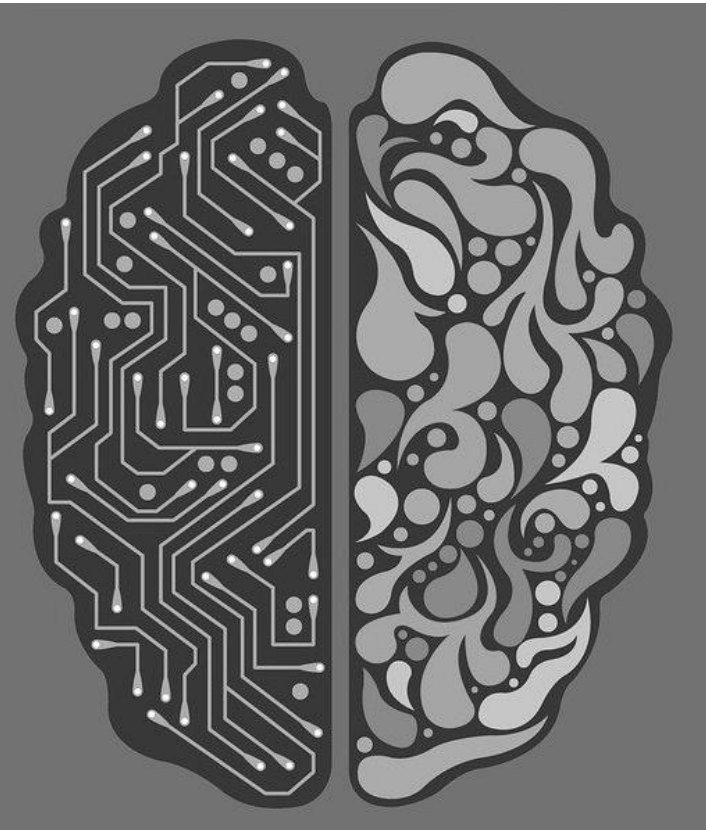
1.3 What is Deep Learning?



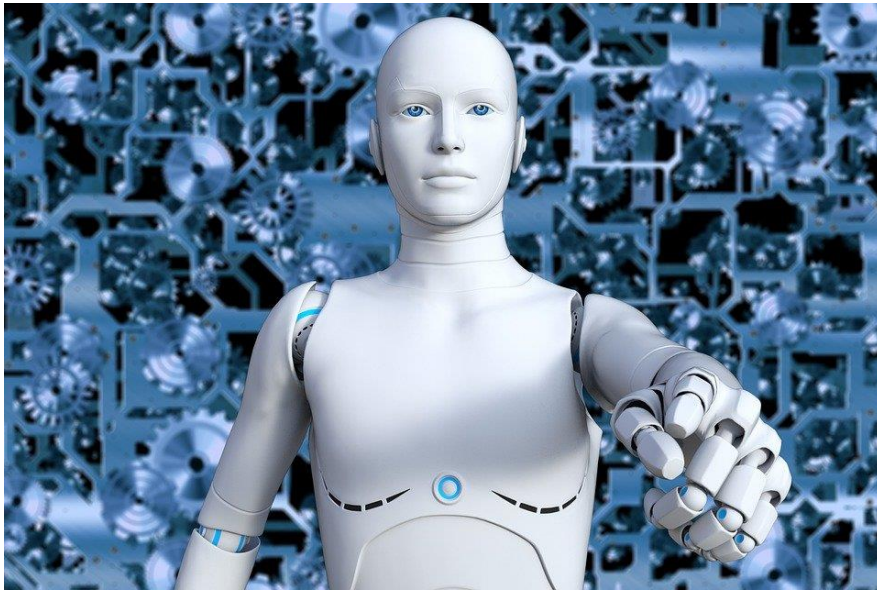
- Deep Learning is a buzz-word!
- The term „deep learning“ was/is a „marketing trick“
- Formerly just known as „multilayer artificial neural networks“ or „multilayer perceptron“



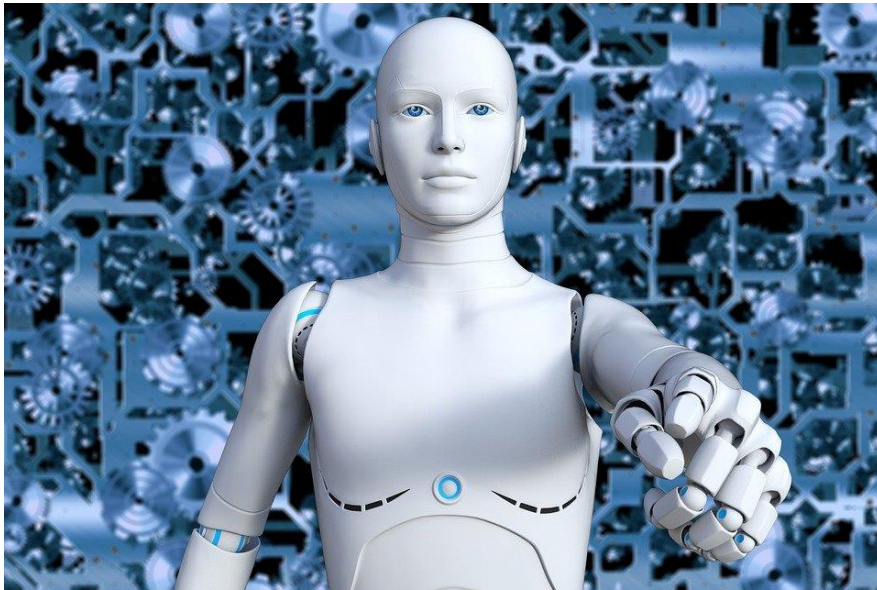
- Artificial Intelligence:
„intelligence exhibited by machines or software”



- ***„Intelligence exhibited by machines or software”***
- ***Intelligence:***
 - *“a general capacity of an individual consciously to adjust his thinking to new requirements” (Stern 1912)*
 - *“Intelligence is, what an intelligence test measures” (Boring 1912)*
- ***„AI is what cannot be done yet by computers”***
(moving goal posts)



...and in (today's) Reality



Strong AI

A computer program (agent), that can understand and solve arbitrary intellectual tasks (that humans can solve)

Weak AI

A computer program that can solve complex tasks in a clearly defined close context

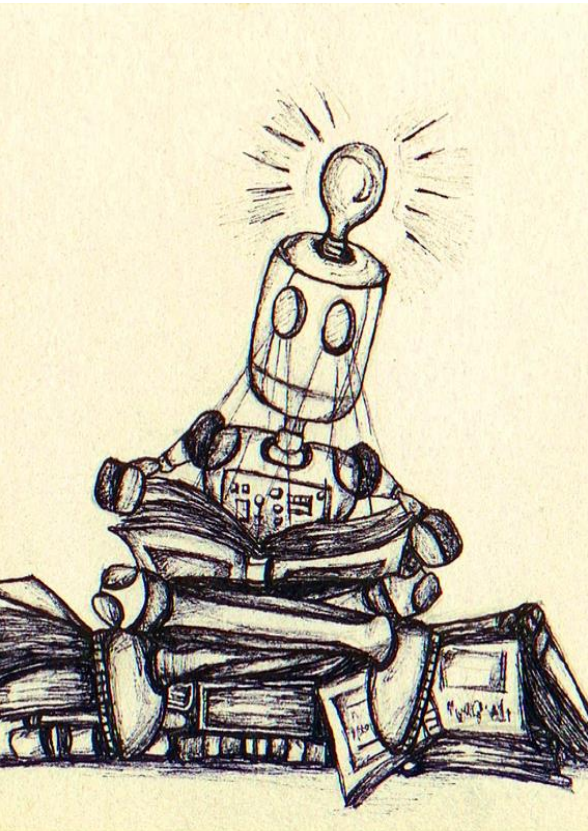


We are here!

- Logic and Reasoning
- Knowledge Representation
- Automated Planning and Scheduling
- **Machine Learning**
- Examples:
 - Solving constraint problems
 - Applying manually created rules
 - ...

“Realistically, deep learning is only part of the larger challenge of building intelligent machines. Such techniques lack ways of representing causal relationship (...) have no obvious ways of performing logical inferences, and they are also still a long way from integrating abstract knowledge, such as information about what objects are, what they are for, and how they are typically used. The most powerful AI systems, like Watson (...) use techniques like deep learning as just one element in a very complicated ensemble of techniques, ranging from the statistical technique of Bayesian inference to deductive reasoning.”

- Gary Marcus



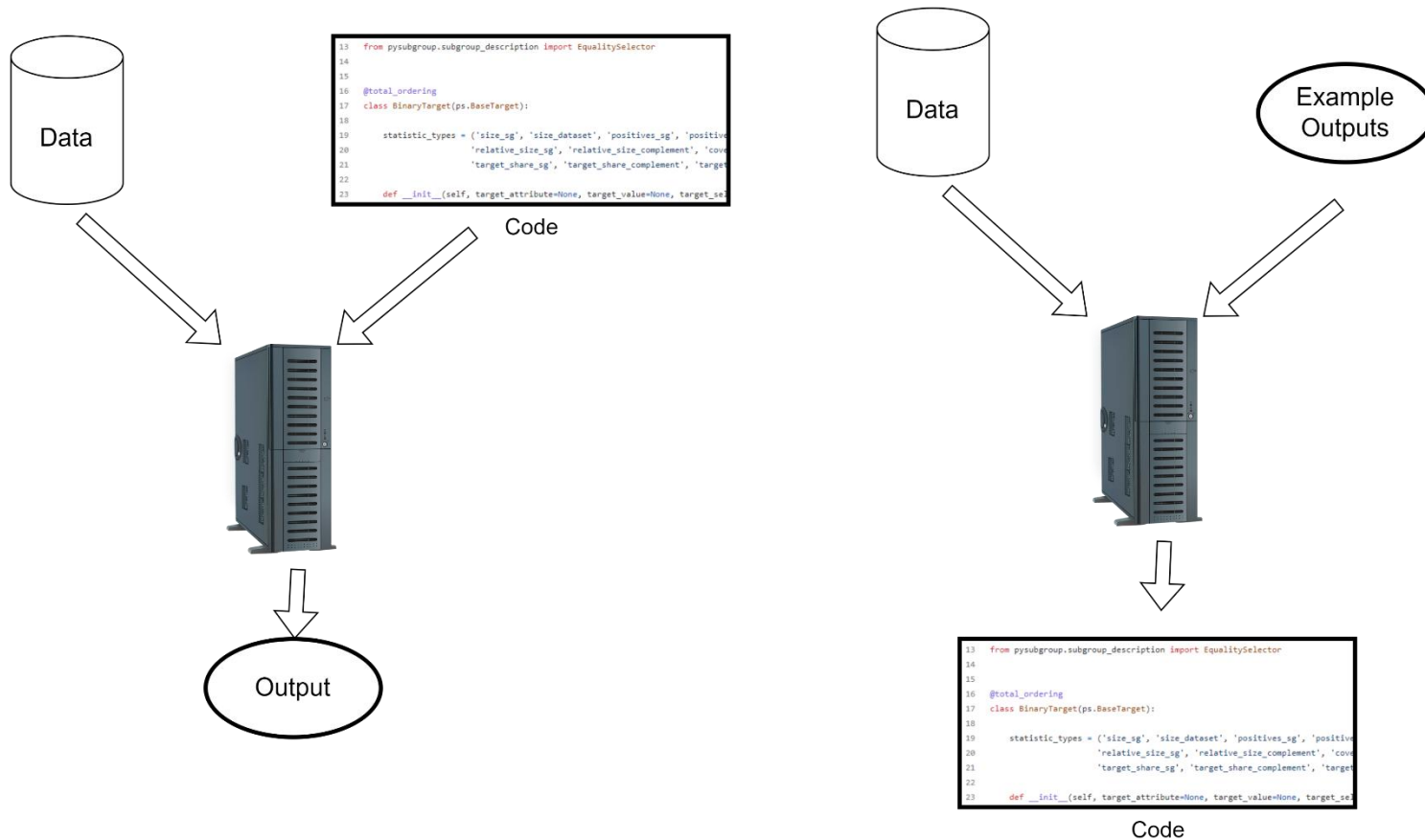
- „**Machine learning (ML)** is the study of computer algorithms that can improve automatically through experience and by the use of data. It is seen as a part of artificial intelligence. Machine learning algorithms build a model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to do so.” - Wikipedia
- Learn a **model** from **training** examples, apply model to make **predictions** about the future

- Classification
- Regression (numeric prediction)
- Structured Prediction
- Clustering
- Reinforcement Learning
- ...

“Deep Learning is a subfield of machine learning concerned with algorithms inspired by the structure and function of the brain called artificial neural networks.”

- Wikipedia

Zhang, W. J., et al. "On definition of deep learning." *2018 World automation congress (WAC)*. IEEE, 2018.

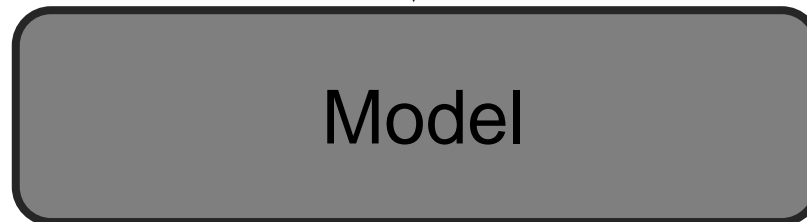


Slide idea: P. Domingos

1001	Windows	Firefox	.com	510	1	Yes
1002	Linux	Chrome	.id	210	145	No
1003	MacOS	Safari	.com	150	1.5	Yes



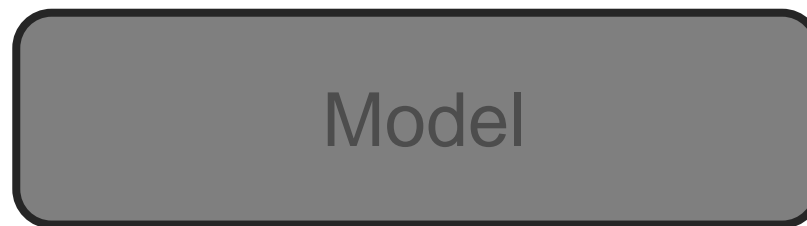
Feed training data into the model to learn



1001	Windows	Firefox	.com	510	1	?
1002	Linux	Chrome	.id	210	145	?
1003	MacOS	Safari	.com	150	1.5	?



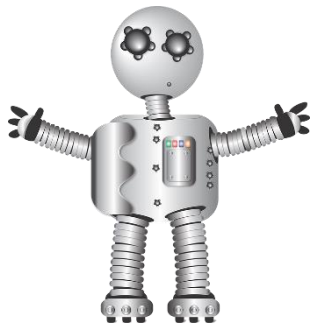
Feed new data into the model to apply



1001	Windows	Firefox	.com	510	1	Yes
1002	Linux	Chrome	.id	210	145	No
1003	MacOS	Safari	.com	150	1.5	Yes



oder



ID	OS	User-Agent Browser	Email-TLD domain	Time for registration (in s)	Posts/Tag	Bot?
1	Windows	Chrome	.de	17	4	Yes
2	Linux	Firefox	.de	412	2	No
3	Windows	Firefox	com	538	0.5	No
4	Linux	Chrome	.fr	392	12	No
5	Windows	Chrome	com	129	1	No
7	Linux	Firefox	com	25	1	No
8	No	Firefox	.es	540	3.5	No
9	Windows	Chrome	.com	58	2	No
10	Linux	Firefox	.de	89	7	No
11	Linux	Chrome	.fr	6	145	Yes
12	Windows	Chrome	.com	572	2	Yes
13	Linux	Safari	.gov	150	5	No
14	Linux	Safari	.id	98	0.3	No
15	Windows	Chrome	.com	159	2	
...
1001	Windows	Firefox	.com	510	1	?
1002	Linux	Chrome	.id	210	145	?
1003	MacOS	Safari	.com	150	1.5	?

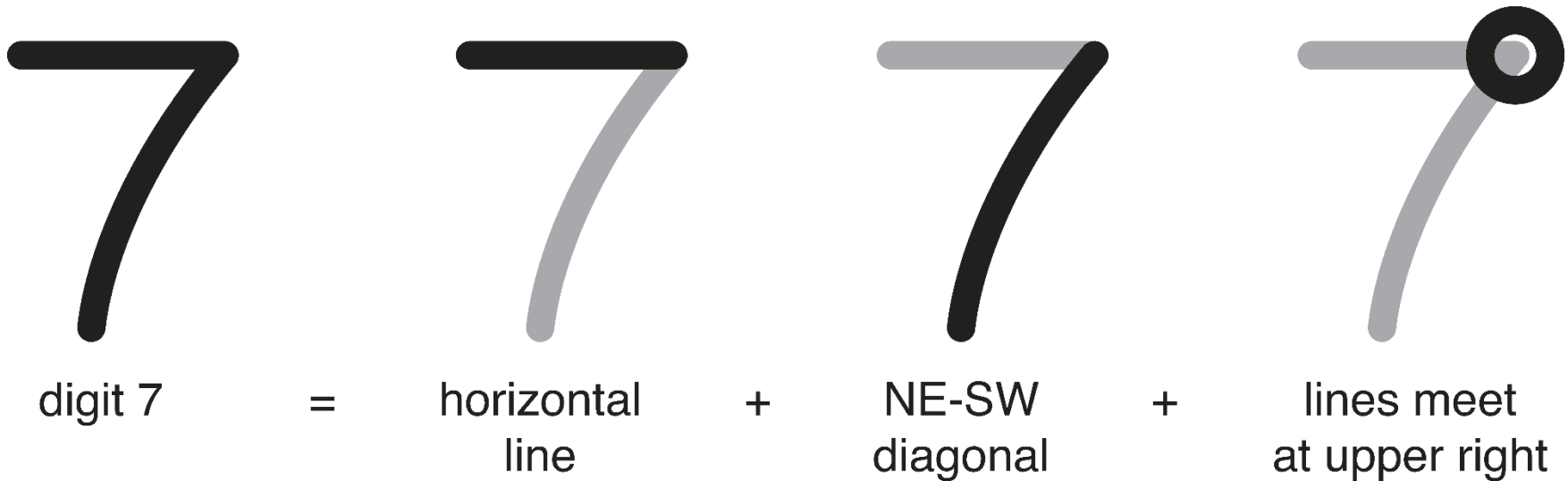
Training

Prediction

- More on Machine Learning: Next Week

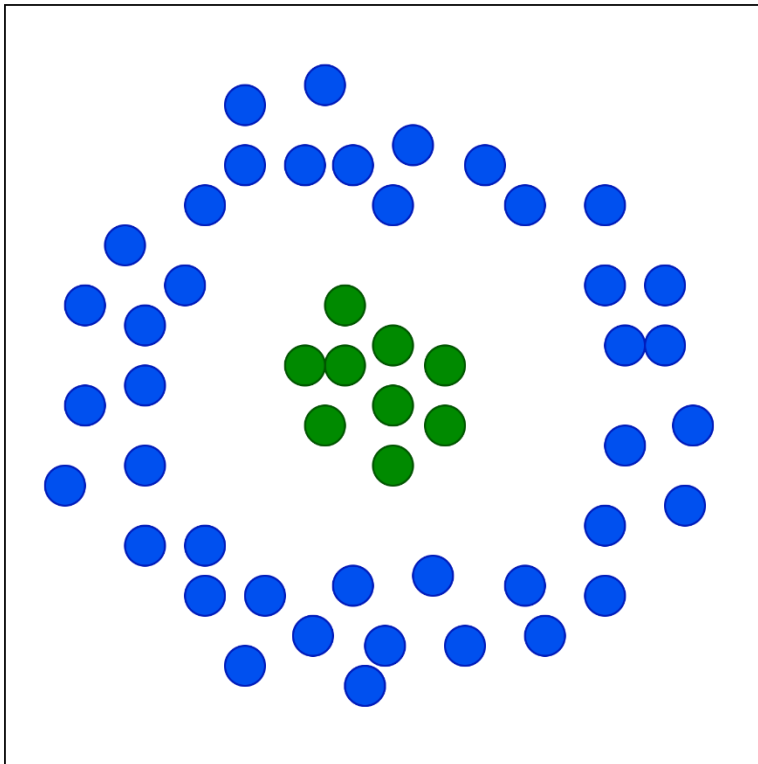
- Machine learning typically uses *features* (properties) of data instances to use in predictions
- Good features are very often not directly contained in the data
- Traditionally, features could get elicited by hand to help prediction
- Features could automatically get created, e.g., identifying rules

Example: Representations

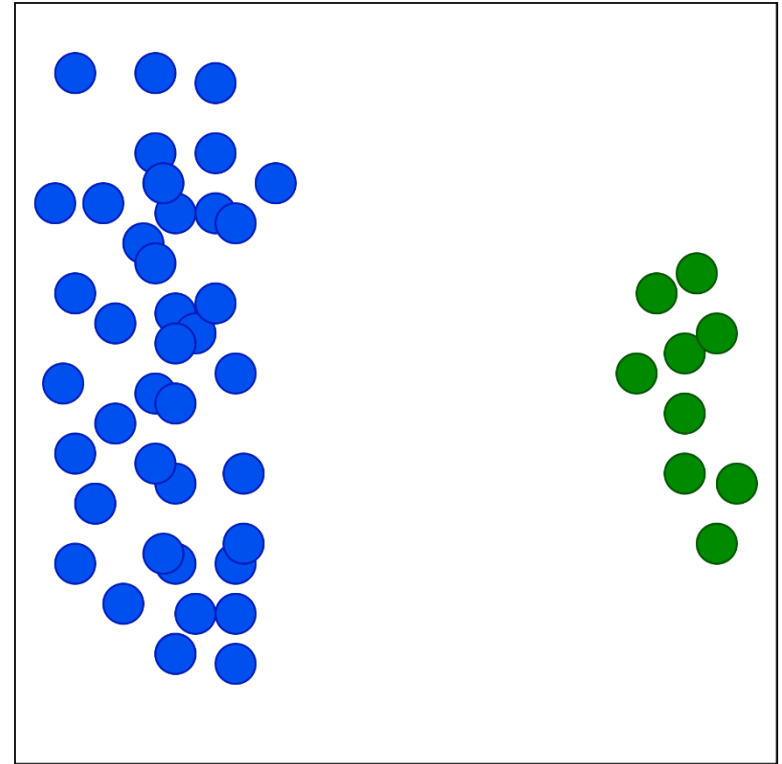


From: "Deep Learning - A Visual Approach" by Andrew Glassner

Cartesian coordinates



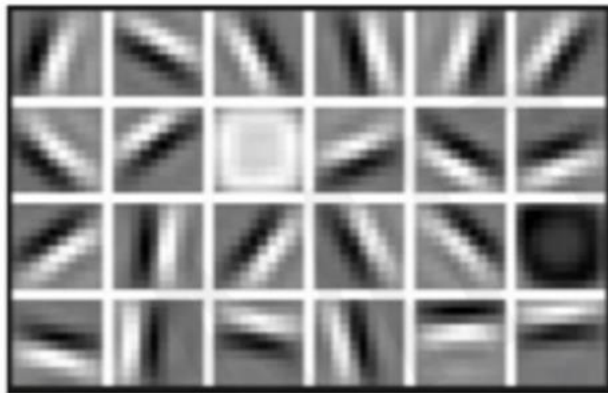
Polar coordinates



Slide idea: https://www.deeplearningbook.org/slides/01_intro.pdf

- Hand-constructed features are time and work consuming
- Often not scalable
- Deep Learning: Construct features automatically from lower level features
- With Artificial Neural Networks

Low Level Features



Lines & Edges

Mid Level Features



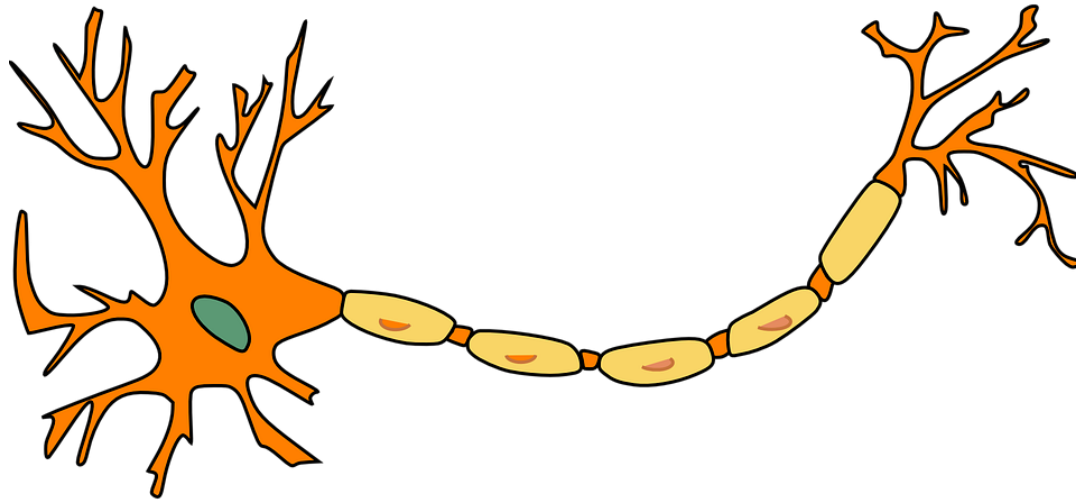
Eyes & Nose & Ears

High Level Features



Facial Structure

© Alexander Amini and Ava Amini, MIT Introduction to Deep Learning
http://introtodeeplearning.com/slides/6S191_MIT_DeepLearning_L1.pdf

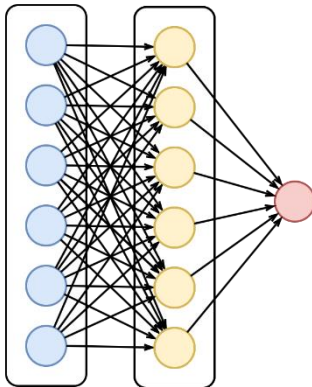


- How?
- Intuition inspired by the biology of the human brain
- Human Neuron
 - Electrically excitable cells
 - Connected to other cells via synapses

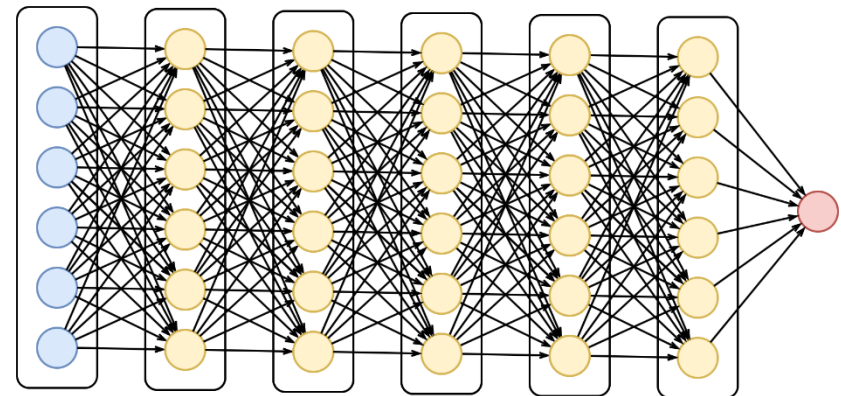
https://en.wikipedia.org/wiki/Artificial_neural_network

Artificial Neural Networks are composite functions

„Shallow“



„Deep“

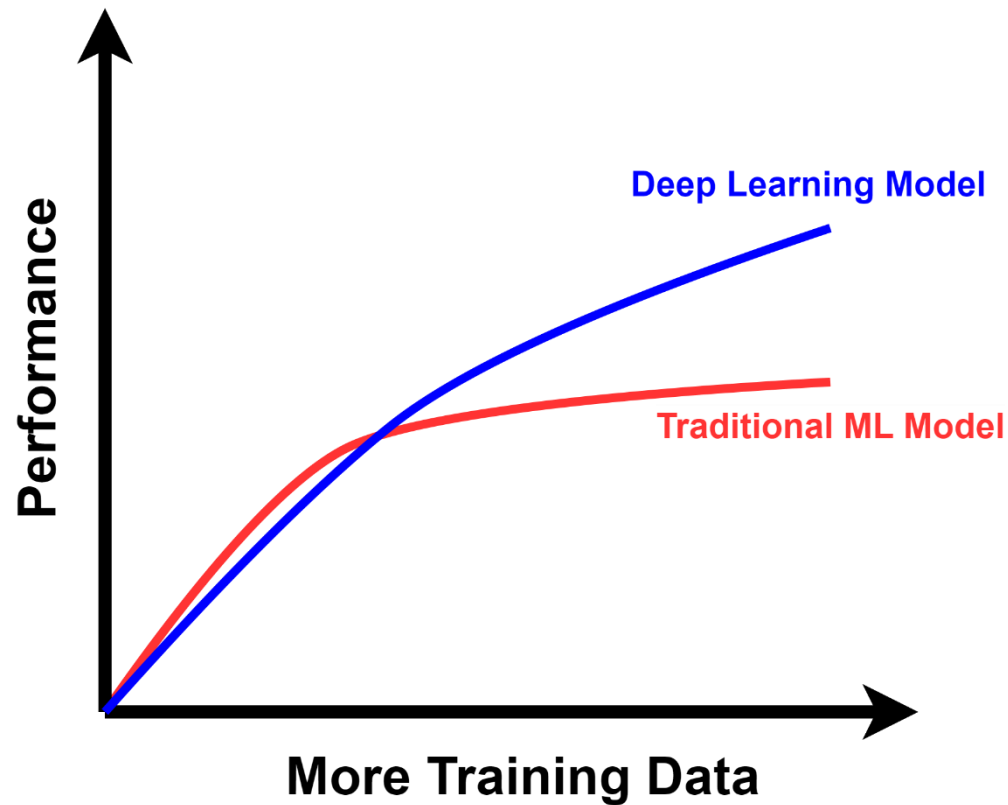


- Each layer in the network is a function that
 - Transforms the last representation
 - Into a new representation of the input data
- Many layer → Cascade of layers building on top of each other

“Deep learning is a class of machine learning algorithms that: (1) use a cascade of multiple layers of nonlinear processing units for feature extraction and transformation. Each successive layer uses the output from the previous layer as input, (2) learn multiple levels of representations that correspond to different levels of abstraction; the levels form a hierarchy of concepts.”

Zhang, W. J., et al. "On definition of deep learning." *2018 World automation congress (WAC)*. IEEE, 2018.

The main claim of Deep Learning



Idea: <https://machinelearningmastery.com/what-is-deep-learning/>

- Deep Learning is much driven by practical applications
- Practical success is often more important than theory and rigor
- More engineering discipline than science?

1.4 A Short History of Deep Learning



- **1943: McCullock and Pitts**

- Early model for neural activation
- Linear threshold neuron (binary):

$$f_{\mathbf{w}}(\mathbf{x}) = \begin{cases} +1 & \text{if } \mathbf{w}^T \mathbf{x} \geq 0 \\ -1 & \text{otherwise} \end{cases}$$

- **1958-1962: Rosenblatts Perceptron**

- First algorithm and implementation to train single linear threshold neuron
- Optimization of perceptron criterion:

- **1979 Fukushima's Neocognitron**

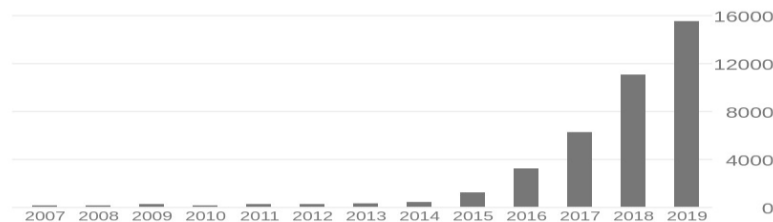
- Multilayer network with different types of cells
- Inspiration for modern CNNs

https://www.youtube.com/watch?v=cNxadbrN_al

- 1969
- Several negative results:
- Famously:
The XOR-problem
Simple perceptrons cannot learn many things
- In the following decade: The „AI Winter“
- Symbolic AI dominates 70s (and 80s)

- 1986
- Efficient calculation of gradients in a deep network wrt. network weights
- Enables application of gradient-based learning to deep networks
- Known since 1961, but first empirical success in 1986
- Remains main workhorse today

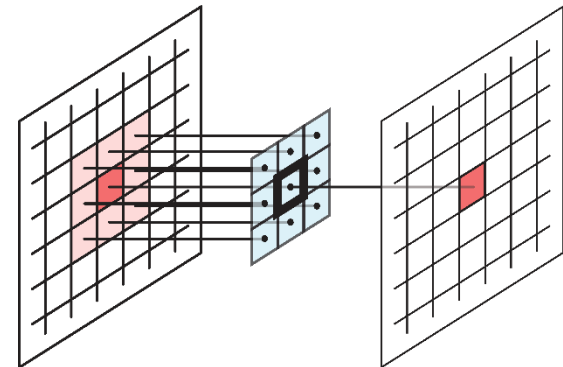
- 1997
- In 1991, Hochreiter demonstrated the problem of vanishing/exploding gradients in his Diploma Thesis
- Led to development of long-short term memory for sequence modeling
- Uses feedback and forget/keep gate



Scholar.google.com citations of the LSTM paper

Summarized from: A. Geiger Deep Learning; Lecture 01 - Introduction

- 1998
- Similar to Neocognitron, but trained end-to-end using backpropagation
- Implements spatial invariance via convolutions and max-pooling
- Weight sharing reduces parameters
- Tanh/Softmax activations
- Good results on MNIST
- But did not scale up (yet)



From: "Deep Learning - A Visual Approach" by Andrew Glassner

- 2006
- Hinton et al.: Training **Deep Belief Networks**
- Deep Learning vs Multilayer neural networks:
 - Intentional Re-Branding
 - No single new inventions
 - Many small improvements and scaling to large data

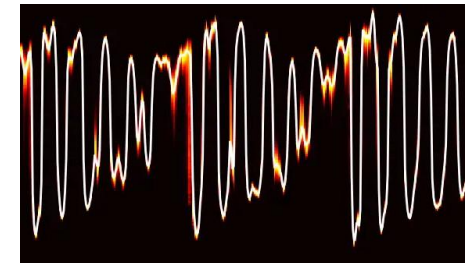
- ImageNet: 2009
 - Large Benchmark Dataset
 - 10,000,000 images
 - 1000 categories
- GPU Computation:
 - 2009 / 2010
 - Raina and Ciresan train Deep Neural Networks with GPUs
 - Massive parallelism enables larger-scale computations

Krizhevsky, Sutskever, Hinton. ImageNet classification with deep convolutional neural networks. NIPS, 2012.

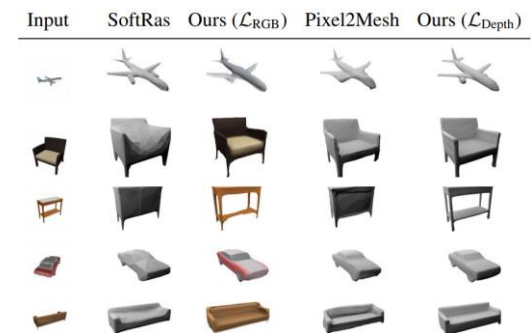
Adapted from: A. Geiger Deep Learning; Lecture 01 - Introduction

- 2013 / 2014
- Word2Vec, Glove
- Models can be (unsupervised) trained based on a general large corpus
- Specialized models then can be obtained by training on smaller training datasets (Fine-Tuning)

- Since 2014 Deep Learning starts to dominate Machine Learning
 - First Images
 - Then Texts
 - Networks / other complex data
- New Techniques and applications (see above)
 - 2016: WaveNet: Generates speech & Music
 - 2016: Style Transfer
 - 2016: AlphaGo
 - 2016 onwards: 3D-Deep Learning



Oord et al.: WaveNet: A Generative Model for Raw Audio. Arxiv, 2016.
Niemeyer, Mescheder, Oechsle, Geiger: Differentiable Volumetric Rendering: Learning Implicit 3D Representations without 3D Supervision. CVPR, 2020.



- 2017 / 2018
- Transformers: Attention replaces recurrence and convolutions
- BERT: Pre-training of language models on unlabeled text
- GLUE: Superhuman performance on some language understanding tasks
- (paraphrase, question answering, ..)
- But: Computers still fail in dialogue

- 2020
- GPT-3
- Language model by OpenAI
- 175 Billion parameters
- Text-in / text-out interface
- Many use cases: coding, poetry, blogging, news articles, chatbots
- Controversial discussions
- Licensed exclusively to Microsoft

- Starts mid 2022
- Currently three main models / competitors
 - Stable Diffusion,
 - Midjourney
 - Dall-E 2



What is ChatGPT?



ChatGPT is a large language model developed by OpenAI, based on the GPT-3.5 architecture. It is an artificial intelligence (AI) model designed to generate human-like text responses in a conversational manner. ChatGPT is trained on a vast amount of text data from the internet, which allows it to generate coherent and contextually relevant responses to a wide range of queries and prompts.

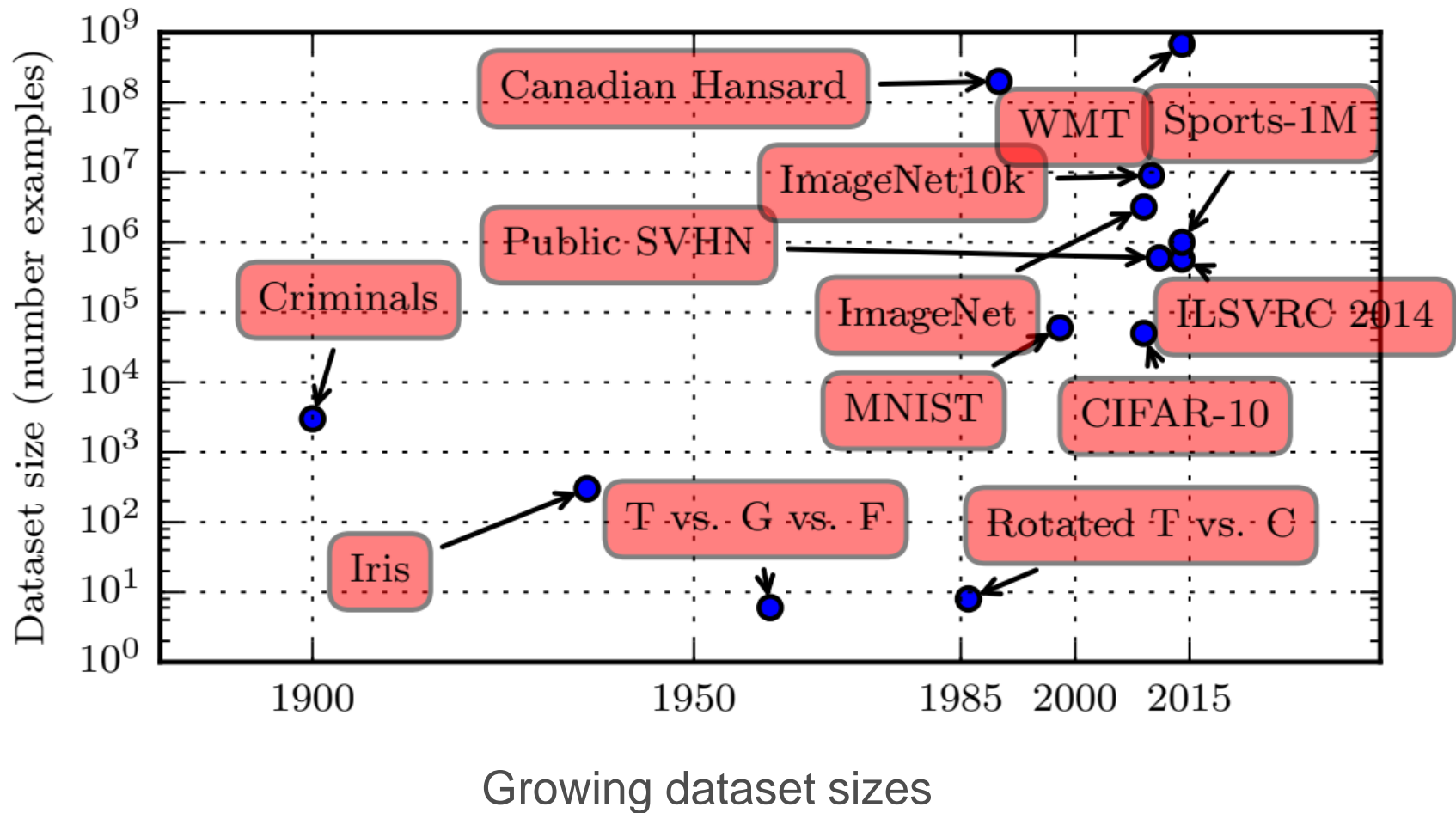


As a language model, ChatGPT can understand and generate text in natural language, which makes it suitable for a variety of applications such as virtual assistants, customer service chatbots, language translation, content generation, and more. ChatGPT is capable of engaging in interactive conversations with users, responding to their prompts, and providing text-based responses in a conversational format. It can generate responses based on its training data and does not have its own real-world experiences or emotions.

pact

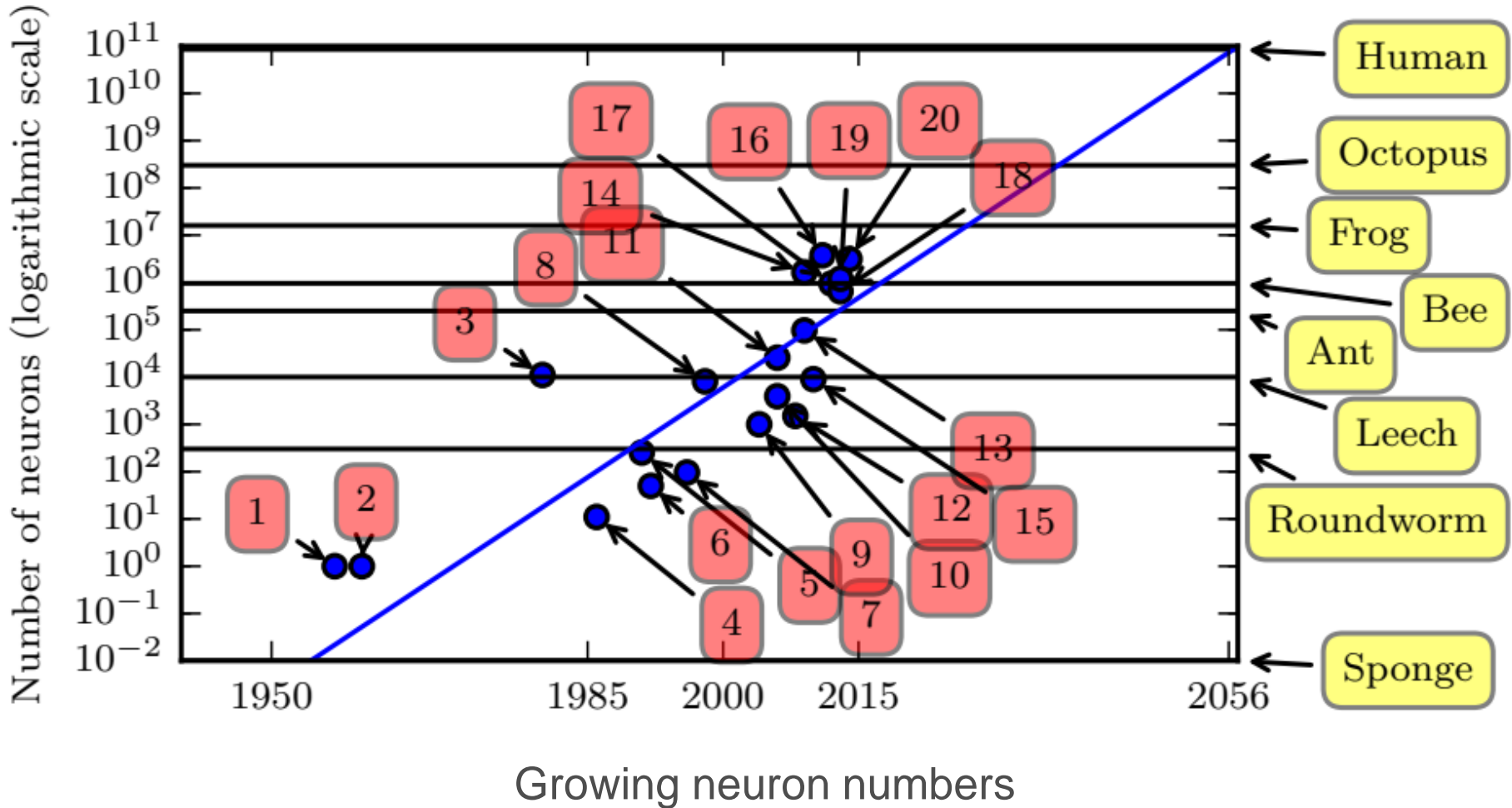
<https://chat.openai.com/>

- Multimodality
- Interpretability
- Privacy
- Vulnerability
- Robustness
- ...



<https://www.deeplearningbook.org/contents/intro.html>

Exponential Growth in Deep Learning



<https://www.deeplearningbook.org/contents/intro.html>