Professur
Künstliche Intelligenz

Introduction into Artificial Intelligence

CHEMNITZ UNIVERSITY OF TECHNOLOGY

Prof. Dr. Fred Hamker
Department of Computer Science
Artificial Intelligence

Introduction

Contents:

What is AI?

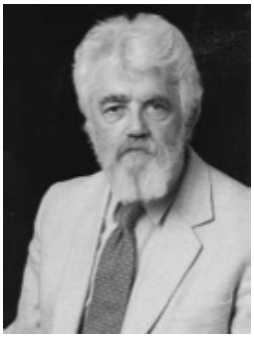
- What is intelligence?
- How can you measure intelligence?
- How does our brain work?

- How can I build a machine that behaves like a human?
- As much as I want to understand the brain or can intelligence also completely different model?

What is AI?

The science and engineering of making intelligent machines

John McCarthy (1955)



John McCarthy (1955)

What is AI?

Elaine Rich (1983): *Artificial intelligence is the study of how to make computers do things at which, at the moment, people are better.*

Intelligenz scheint eine Eigenschaft von Menschen zu sein, die sich in einer unvorhergsehenen Situation zurechtfinden können. Das ist etwas anderes, weshalb man den Begriff "intelligent" für Maschinen besser vermeiden sollte.

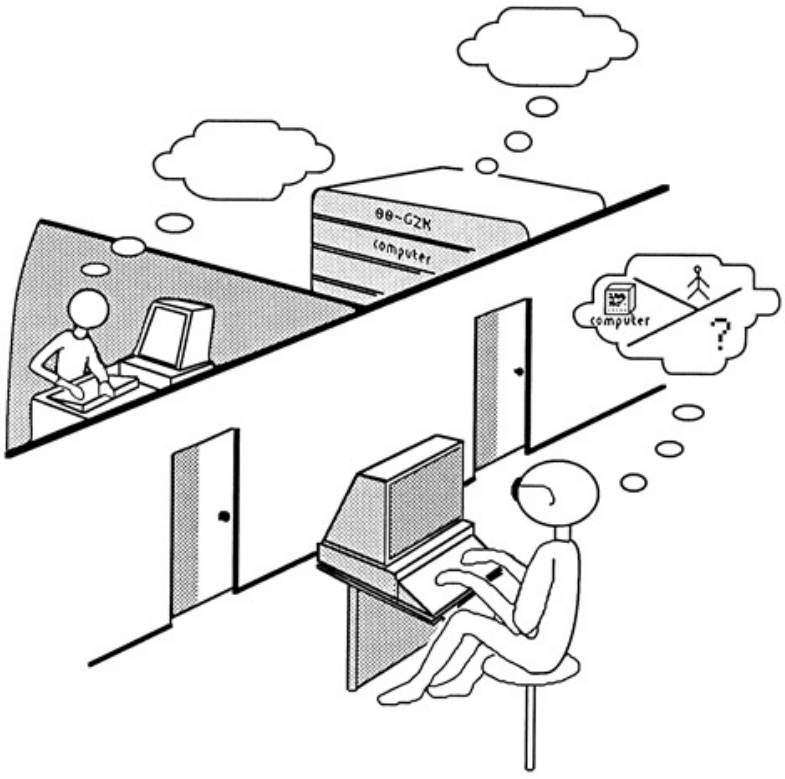
Peter Rechenberg, Informatik Spektrum 33(1), 2010

What is AI?

Strong KI:
Create intelligence that think like humans and can solve problems and is characterized by a form of consciousness or self-awareness and emotions.

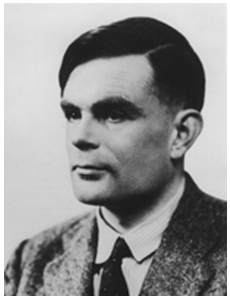
Weak KI:
Applications of interest, seems to be necessary to the solution of a form of "intelligence". Simulation of intelligent behavior by means of mathematics and computer science. It's not about creating awareness and a deeper understanding of intelligence.

The Turing Test



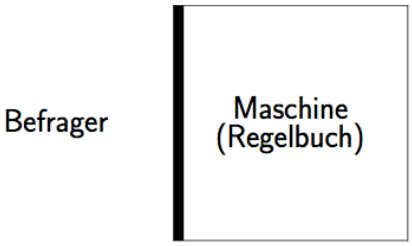
The test person Alice is sitting in a locked room with two computer terminals.
A terminal is connected to the machine, the other with the person Bob.

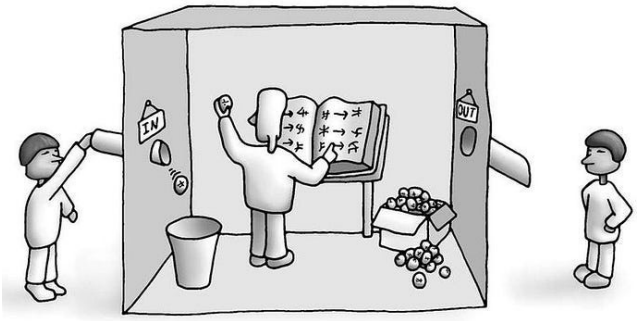
Alice can ask questions. It has the task of deciding after 5 minutes, which terminal the machine answers. The machine passes the test if they can fool Alice in 30% of cases.



Alan Turing (1912-1954)

Criticism of Searle at the Turing test: Chinese Room



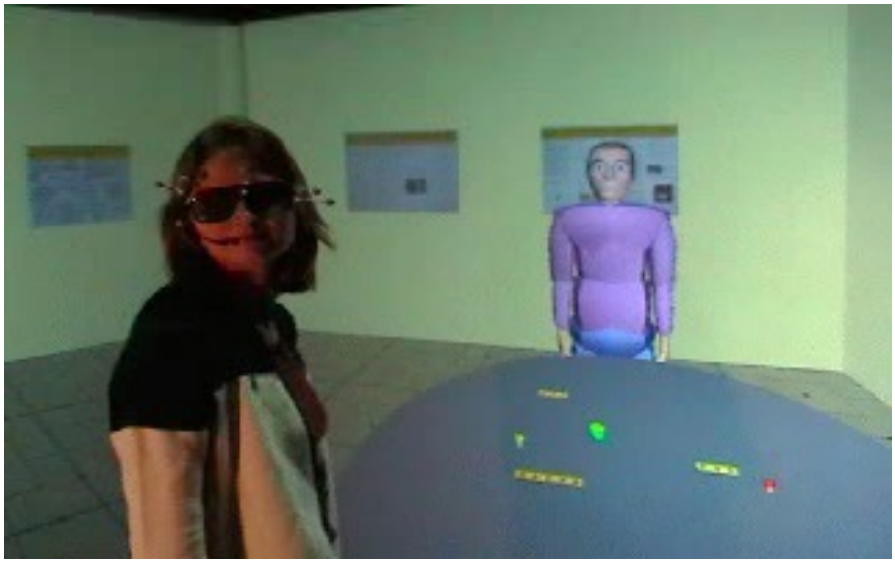


Imagine that you do not speak Chinese, but have a huge rule book that allows you to look up Chinese sentences and afterwards tells you how to respond in Chinese. Chinese you do not understand, can behave in a seemingly intelligent manner, however.

> Searle argued that computers, even if they were appearing intelligent, it is not real, because they were just using something like the rule book of the Chinese room. You do not really understand the meaning of the content.

Example Affective Computing

There is no machine, that has real emotions, but they make us believe they have emotions.



The Turing Test

The Turing test defines the following disciplines working within the AI:

1. Knowledge representation: to store what it knows or hears
2. Automated reasoning: to use stored information to answer questions and to draw new conclusions
3. Natural language processing (written language): to enable it to communicate successfully in English
4. Machine learning: to adapt to new circumstances and to detect and extrapolate pattern

Turing's test deliberately avoided direct physical interaction between the interrogator and the computer, because physical simulation of a person is unnecessary for intelligence.

Total Turing-Test: Capture full term intelligence

5. Computer vision: to perceive objects
6. Understanding and synthesis of speech
7. Robotics: to manipulate objects and move about

These seven disciplines form the largest part of the AI

Goals of AI

Long-term goal:
To build systems that meet or exceed the performance of humans on a large scale. Whether this goal in principle be reached (and ethically / economically desirable), it is has long been hotly debated.

Short-term goal:
To build specialized systems that can solve certain tasks.

The History of AI

- Early influences
  - Gödel (1931) Completeness of first-order predicate logic
- The gestation of AI (1943 – 1955)
  - Function of neurons in the brain (McCulloch & Pitts, 1943)
  - Hebbian learning rule of neurons (Hebb 1949)
  - Alan Turing (1950)
  - Minsky (1951) neural network computer: 40 neurons with 3000 vacuum tubes
- The birth of AI, Workshop at Dartmouth College (1956)
  - McCarthy: LISP, Programming language that was created specifically for the processing of symbolic structures
  - Newell & Simon: Logic Theorist: first automatic theorem prover; Showed that computer, otherwise the expected numbers, can process symbols.

The History of AI

- Early enthusiasm, great expectations (1952 – 1969)
  - Newell & Simon (1961) General Problem Solver (GPS) – A program that simulates human thought
  - Samuel (1959) Checkers program
  - Robinson (1965) resolution method
  - Weizenbaum (1966) Eliza: Responds like a human psychologist (
  - Blocks world
  - Perceptron
  - search methods
  - Expectations have been greatly hyped
  - little or no knowledge about the scope

The History of AI

- A dose of reality (1966 – 1973)
  - Systems failed in more difficult tasks (machine translation)
  - Complexity of many problems
  - Minsky & Papert (1969) The perceptron can xor problem not solve
- Knowledge-based systems (expert systems) (1969 – 1979)
  - de Dombal (1972) Expert system for diagnosis of abdominal diseases
  - Colmerauer (1972) logic programming language PROLOG
  - Shortliffe & Buchanan (1976) MYCIN: medical expert system (about 450 rules, can catch up with some experts; better than young doctors)

The History of AI

- AI becomes an industry (1980 - present)
  - The first successful commercial expert systems
  - DART (Dynamic Analysis and Replanning Tool), automated logistics planning and schedules for transportation tasks.
  - Japan starts with great effort the "5th generation project". Objective: To build Efficient PROLOG machines.
  - Big Data (Google, Facebook, Microsoft)
  - Large Language Model (LLM)
- The return of neural networks (1986 - 2000)
  - Disappointment about the possibilities of expert systems
  - Rumelhart, Hinton & Williams (1986) Backpropagation
  - Kohonen (1995) Self-Organizing Feature Maps (SOM)
  - Fukushima (1980) Neokognitron, NN for object recognition
  - Convolutional Networks (LeCun, 1990)

The History of AI

- AI becomes a science (1987 - present)
  - Hidden Markov Models
  - Pearl (1988) Bayesian networks
  - Vapnik (1995) Support vector machine – Classifier based on a statistical learning theory
  - Barto & Sutton (1998) Reinforcement Learning, Learning of agents that are in interaction with the environment through non-specific feedback
- Reinforcement Learning (1990 - present)
  - Learning of agents, that interact with their environment by unspecific feedback (Barto & Sutton, 1998)
- Computational Neuroscience (1995 - present)
  - Disappointment about the possibilities of classical neural networks, in particular as regards the declaration of human capabilities
  - Advent of methods to simulate neural networks bio-inspired
  - currently geared more biologically, but approaches in robotics

The History of AI

- The emergence of intelligent agents (1995 - present)
  - Robots, sensors, modeling of probabilities
  - RoboCup (Soccer)
  - Urban Challenge (Darpa USA)
  - Autonomous driving

The History of AI

- Deep Learning Architectures (2010 until today)
  - Deep Learning networks win all competitions in pattern recognition (IJCNN 2011 Traffic Sign Recognition Competition, ICPR 2012 Mitosis Detection in Breast Cancer Histological Images, ImageNet 2012 Large Scale Visual Recognition Challenge).
  - In 2014 Google buys Deepmind for 400 M€ .
  - Deep RL Learning, from the image to robot control in a single architecture.
  - GPT, ChatGPT: GPT (**G**enerative **P**re-trained **T**ransformer) are large language models, that can learn statistical dependencies of words or parts in sentences and therefore allow for a complex communication and writing of text, images and videos.

Example: Deep RL Learning, from the image to robot control in a single architecture

To what extent do we want Artificial Intelligence

- People could lose their jobs due to the increasing automation (present discussion about Industry 4.0).
- People could lose the appreciation, to be unique.
- People could lose some of their privacy rights.
- The use of AI systems could lead to a loss of responsibility (autonomous driving).
- First victims by autonomous cars (Tesla, 2016, 2018, Uber, 2018)

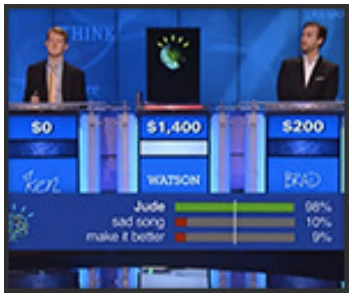
- The success of the AI could mean the end of the human race. Almost every technology has the potential to cause harm in the wrong hands (AI controlled weapons).

A chess computer wins against the chess world champion Garry Kasparov - May 1997

1:0
0:1
½:½
½:½
½:½
0:1

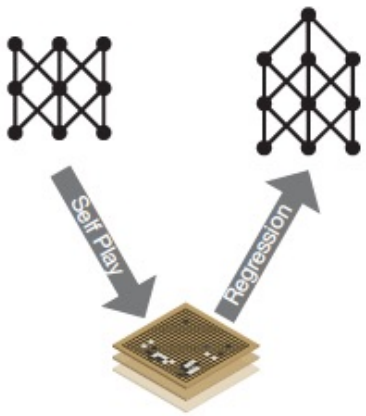
Thus, the system was the first computer that could decide a match under tournament conditions against a reigning world chess champion for themselves.

IBM supercomputer Watson wins Jeopardy Quiz Show - February 2011



- Task: To a note formulate a question. Often cryptic notice (usually formulated deliberately ambiguous)
- Combination of several facts required
- Points can only be obtained who is faster than its two competitors.
- compared to Jeopardy (http://www.jeopardy.com/) is chess a mathematically easy to define game

AlphaGo wins against a Go Grandmaster Lee Sedol – March 2016



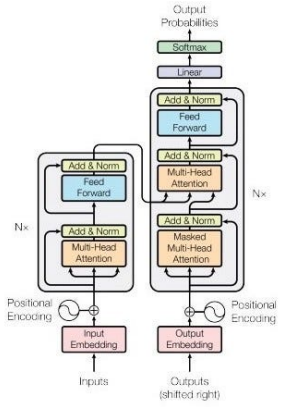
Go has been considered a great challenge for AI. Due to the larger number of possible moves the search trees become very large and impossible to be searched exhaustively. Further, the evaluation of a state of the game is much more difficult than in chess.

AlphaGo wins against a Go Grandmaster Lee Sedol – March 2016

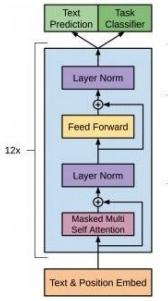
- Combines Monte Carlo Tree Search (MCTS) with learning Neuronal Networks and Reinforcement Learning
- MCTS: Method for building a weighted search tree (fast)
- Policy Network: Determines the action
- Value Network: Estimates the value of a state
- Supervised learning of policy networks (expert knowledge)
- Reinforcement Learning of policy networks (Self-play)
- Reinforcement Learning of value networks
- Combination of MCTS, policy und value networks
- Hardware (Version in Nature: 1202 CPUs und 176 GPUs)

ChatGPT

GPT (Generative Pre-trained Transformer) is a language model (neuronal network) which has been developed by OpenAI on basis of transformer neuronal networks.

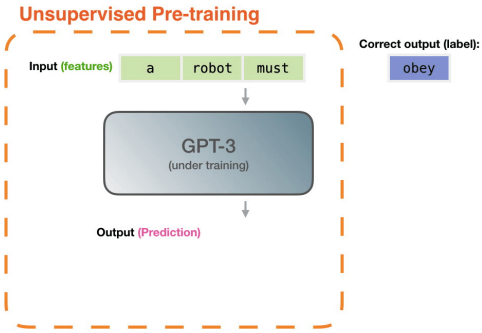


Transformer



GPT

„Human-like conversation“, which has been trained by “next-word prediction“.



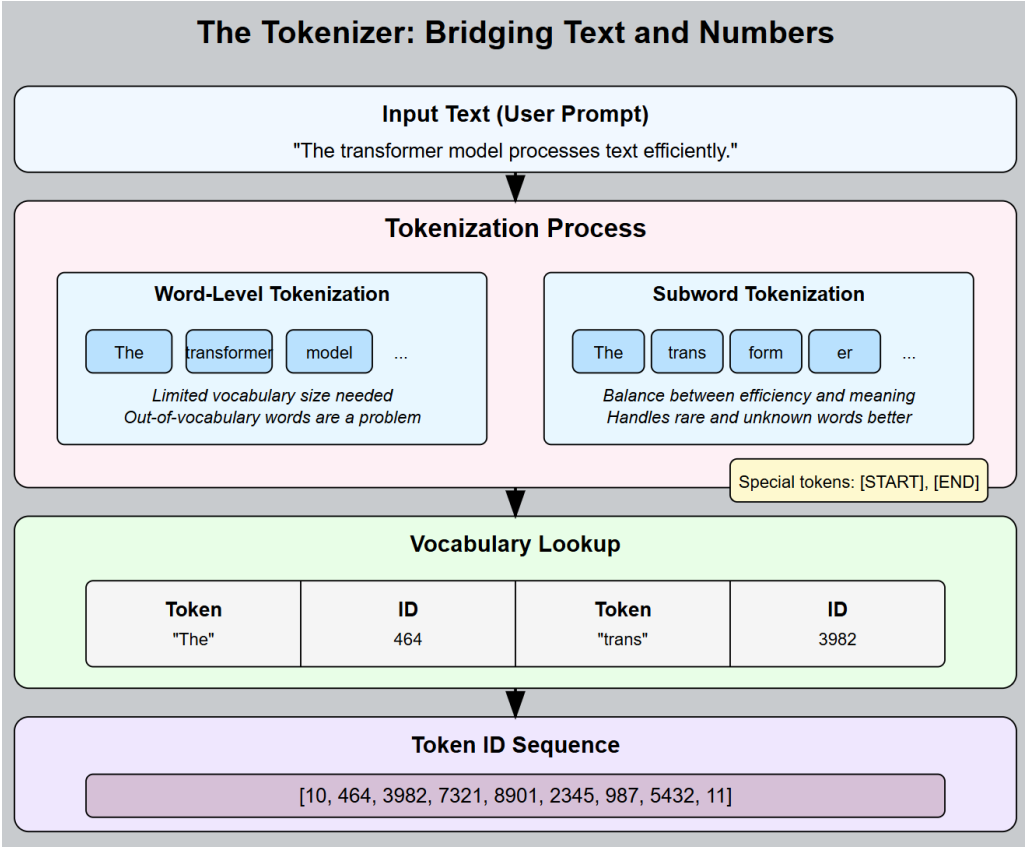
Revolution or just a bullshit Parrrot ?

https://www.reasonfieldlab.com/post/chatgpt-the-revolutionary-bullshit-parrot

Chat GPT has been optimised by supervised und reinforcement learning methods.

Quelle: https://medium.com/nerd-for-tech/gpt3-and-chat-gpt-detailed-architecture-study-deep-nlp-horse-db3af9de8a5d

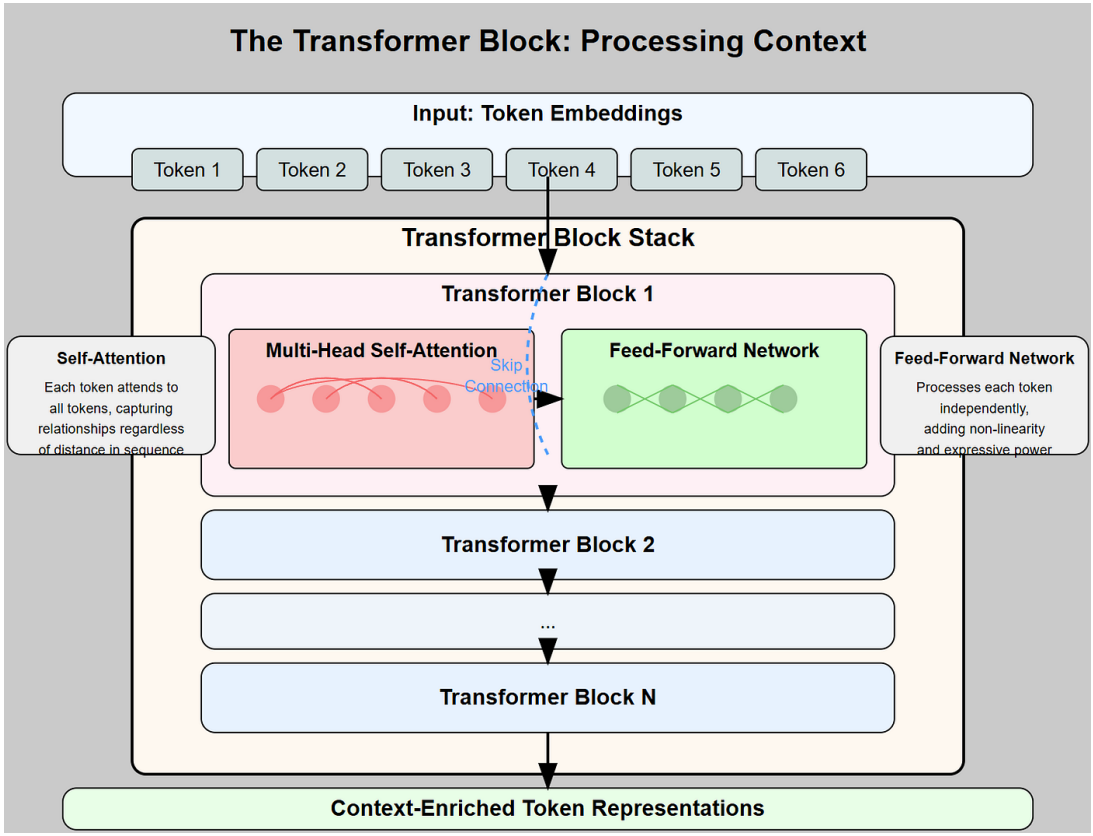
Large Language Models



https://medium.com/codex/how-transformer-llms-generate-text-one-token-at-a-time-5531838bc2a1

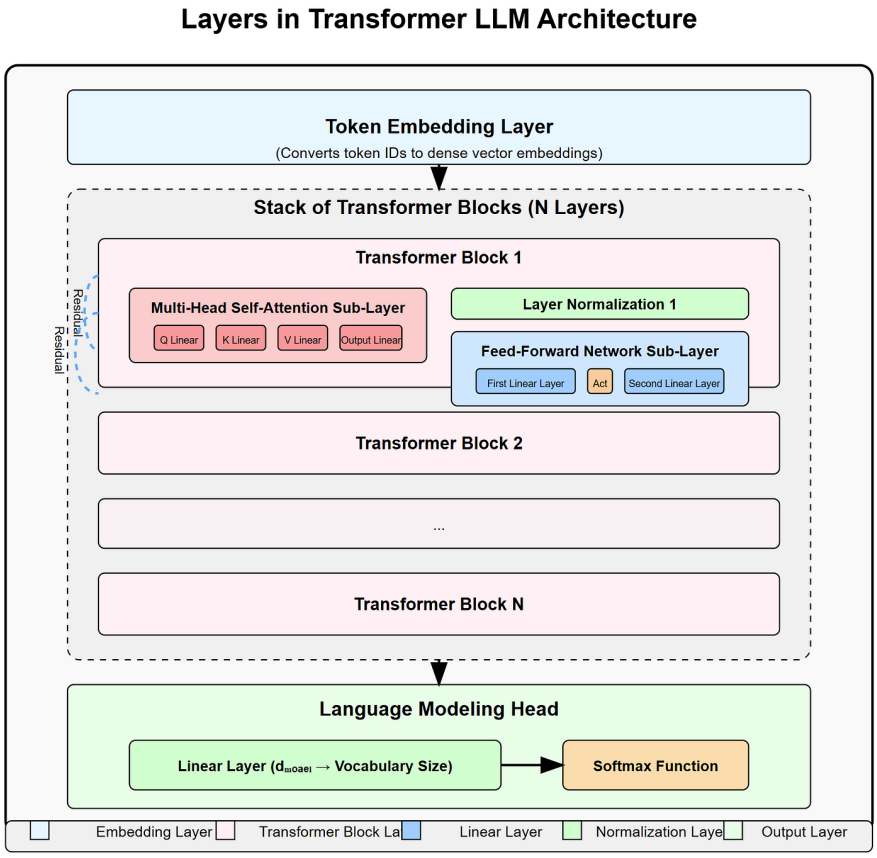
Large Language Models

knowledge is distributed and emergent across the entire set of parameters

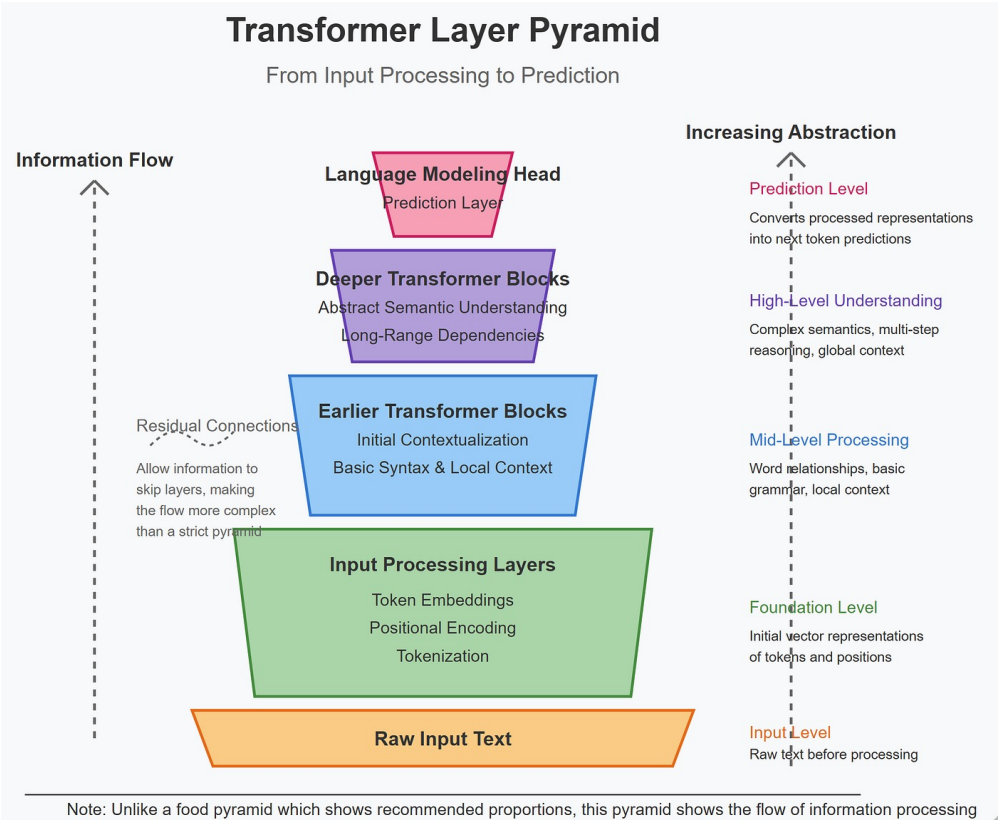


https://medium.com/codex/how-transformer-llms-generate-text-one-token-at-a-time-5531838bc2a1

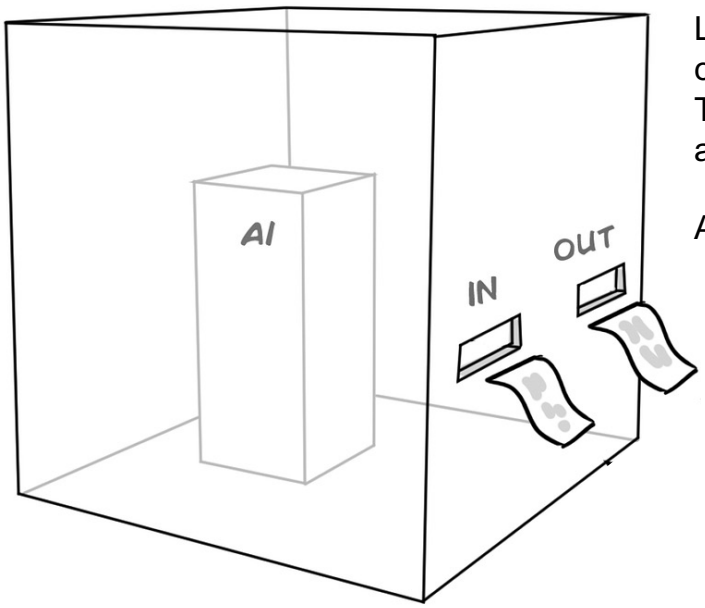
Large Language Models



Large Language Models



Large Language Modelle and Chinese Room



Large Language Models communicate solely via text (tokens). They are trained to predict the next or a masked word.

Are they intelligent?

How far we have reached the goal of being able to build intelligent machines?

- A computer can calculate formulas much better than humans.
- Today's computers, however, are (completely) overwhelmed with the interpretation of a scene and make decisions in the real world.
- People can learn much better than computers

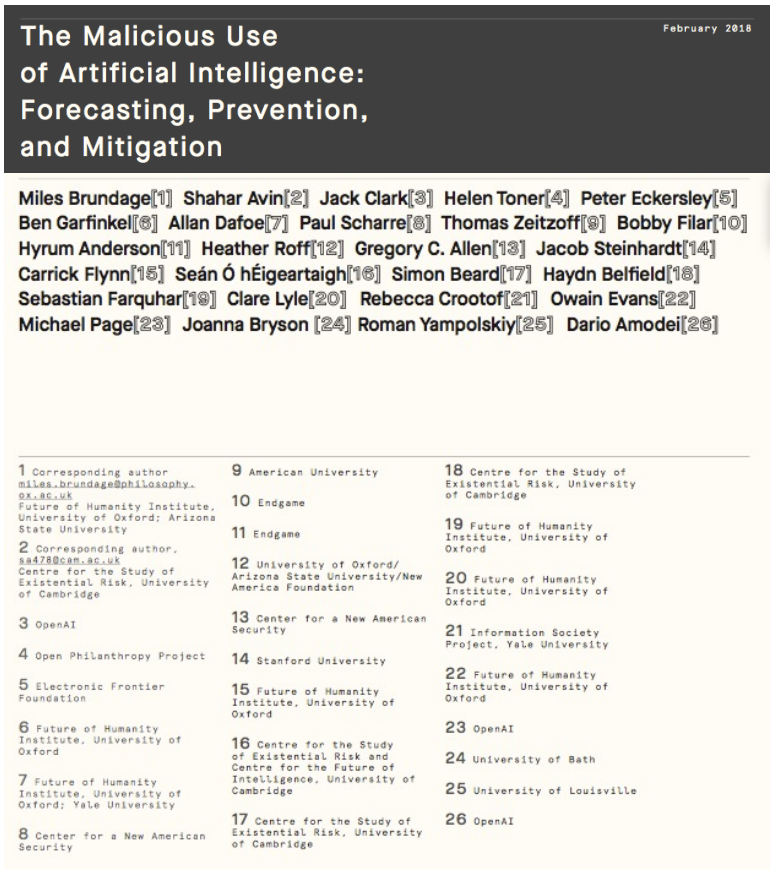
Which do disciplines affect the AI?

- Philosophy (428 v. Chr. - present)
- Mathematics (ca. 800 - present)
  - predicate logic
  - algorithm
  - probability theory
  - game theory
  - Markov decision processes

Which do disciplines affect the AI?

- Neuroscience (1861 - present)
  - How do brains process information?
  - Single-cell recordings, picture added method
  - Computational Neuroscience
- Psychology (1879 - present)
  - How do humans and animals think and act?
  - Cognition
  - patient studies
- Computer engineering (1940 - present)
  - How can we build an efficient computer or robot.
- Control theory and Cybernetics (1948 - present)
- Linguistics (1957 - present)
  - Natural language processing

To what extent do we want Artificial Intelligence



The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation

February 2018

Miles Brundage[1] Shahar Avin[2] Jack Clark[3] Helen Toner[4] Peter Eckersley[5] Ben Garfinkel[6] Allan Dafoe[7] Paul Scharre[8] Thomas Zeitzoff[9] Bobby Filar[10] Hyrum Anderson[11] Heather Roff[12] Gregory C. Allen[13] Jacob Steinhardt[14] Carrick Flynn[15] Seán Ó hÉigeartaigh[16] Simon Beard[17] Haydn Belfield[18] Sebastian Farquhar[19] Clare Lyle[20] Rebecca Crootof[21] Owain Evans[22] Michael Page[23] Joanna Bryson [24] Roman Yampolskiy[25] Dario Amodei[26]

1 Corresponding author miles.brundage@philosophy.ox.ac.uk Future of Humanity Institute, University of Oxford; Arizona State University
2 Corresponding author, sa478@cam.ac.uk Centre for the Study of Existential Risk, University of Cambridge
3 OpenAI
4 Open Philanthropy Project
5 Electronic Frontier Foundation
6 Future of Humanity Institute, University of Oxford
7 Future of Humanity Institute, University of Oxford; Yale University
8 Center for a New American Security
9 American University
10 Endgame
11 Endgame
12 University of Oxford/ Arizona State University/New America Foundation
13 Center for a New American Security
14 Stanford University
15 Future of Humanity Institute, University of Oxford
16 Centre for the Study of Existential Risk and Centre for the Future of Intelligence, University of Cambridge
17 Centre for the Study of Existential Risk, University of Cambridge
18 Centre for the Study of Existential Risk, University of Cambridge
19 Future of Humanity Institute, University of Oxford
20 Future of Humanity Institute, University of Oxford
21 Information Society Project, Yale University
22 Future of Humanity Institute, University of Oxford
23 OpenAI
24 University of Bath
25 University of Louisville
26 OpenAI

Active AI researchers thought about the near development of AI and potential risks of AI misuse.

Among the proposed solutions the also suggested to centralize solutions and not to publish the technology until risks have not been solved

https://arxiv.org/pdf/1802.07228.pdf

To what extent do we want Artificial Intelligence

Pause Giant AI Experiments: An Open Letter

We call on all AI labs to immediately pause for at least 6 months the training of AI systems more powerful than GPT-4.

Signatures
1859

Add your signature

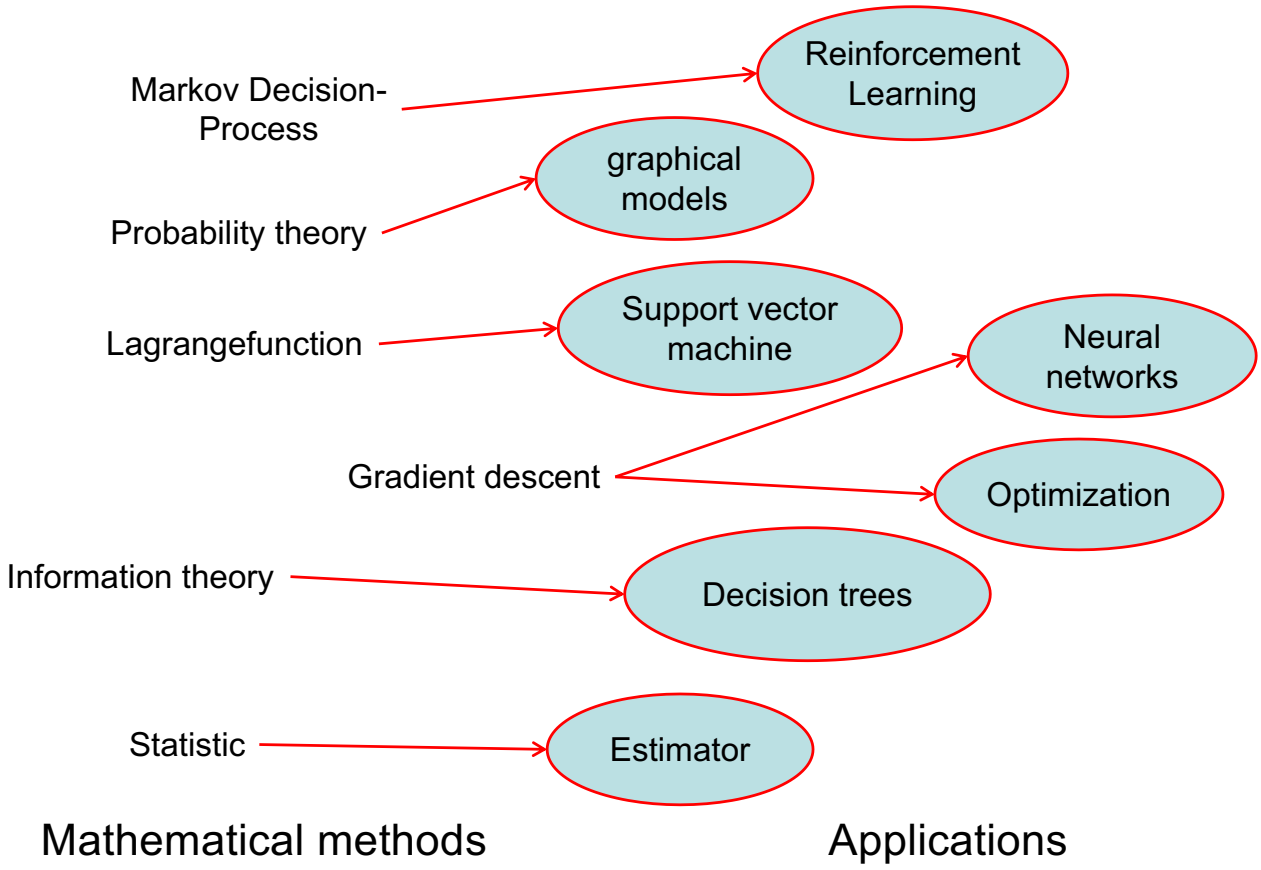
future of life INSTITUTE

„Contemporary AI systems are now becoming human-competitive at general tasks, …: Should we let machines flood our information channels with propaganda and untruth? Should we automate away all the jobs, including the fulfilling ones? Should we develop nonhuman minds that might eventually outnumber, outsmart, obsolete and replace us? Should we risk loss of control of our civilization? Such decisions must not be delegated to unelected tech leaders.
Powerful AI systems should be developed only once we are confident that their effects will be positive and their risks will be manageable. …

Therefore, we call on all AI labs to immediately pause for at least 6 months the training of AI systems more powerful than GPT-4.“

Probably there has been no pause at all …

Overview – Introduction in the AI



Overview – Introduction in the AI

Contents:

- Terms and Concepts
- Search, Games and Problem Solving
- Constraint problem
- Planning (brief overview)
- Optimization
- Neural Networks, Support Vector Machine
- Probability theory
- Information theory
- Decision trees
- Statistic and Estimator
- Reinforcement Lerning
- Probabilistic Graphical Models (Bayesian Networks)

AI at the Chemnitz University of Technology

Einführung in die Künstliche Intelligenz I

Einführung in die Künstliche Intelligenz II

Artificial Intelligence for mobile robots

Bildverstehen

Neuro-Robotik

CHEMNITZ UNIVERSITY OF TECHNOLOGY

The Wide Landscape of Neuromorphic Computing

Neurokognition I + II

Prädiktive Verhaltensanalyse

Neuro-computing

Maschinelles Lernen

Deep Reinforcement Learning

Siehe auch Studiengang Master Neurorobotik

Research in AI on TU Chemnitz

Computational neuroscience

Perception

- Attention
- Body schema, self-awareness
- Category learning
- Cognitive control of visual perception
- Object recognition
- Learning in the visual cortex
- Masking and conscious perception
- Peri-saccadic space perception

Cognition and action

- Basal ganglia, action selection and reinforcement learning
- Hippocampus, episodic memory and spatial navigation
- Cerebellum, forward and inverse models

Neuro-cognitive agents

- Neurorobotics
- Neuronal agents in virtual reality
- Teaming agents

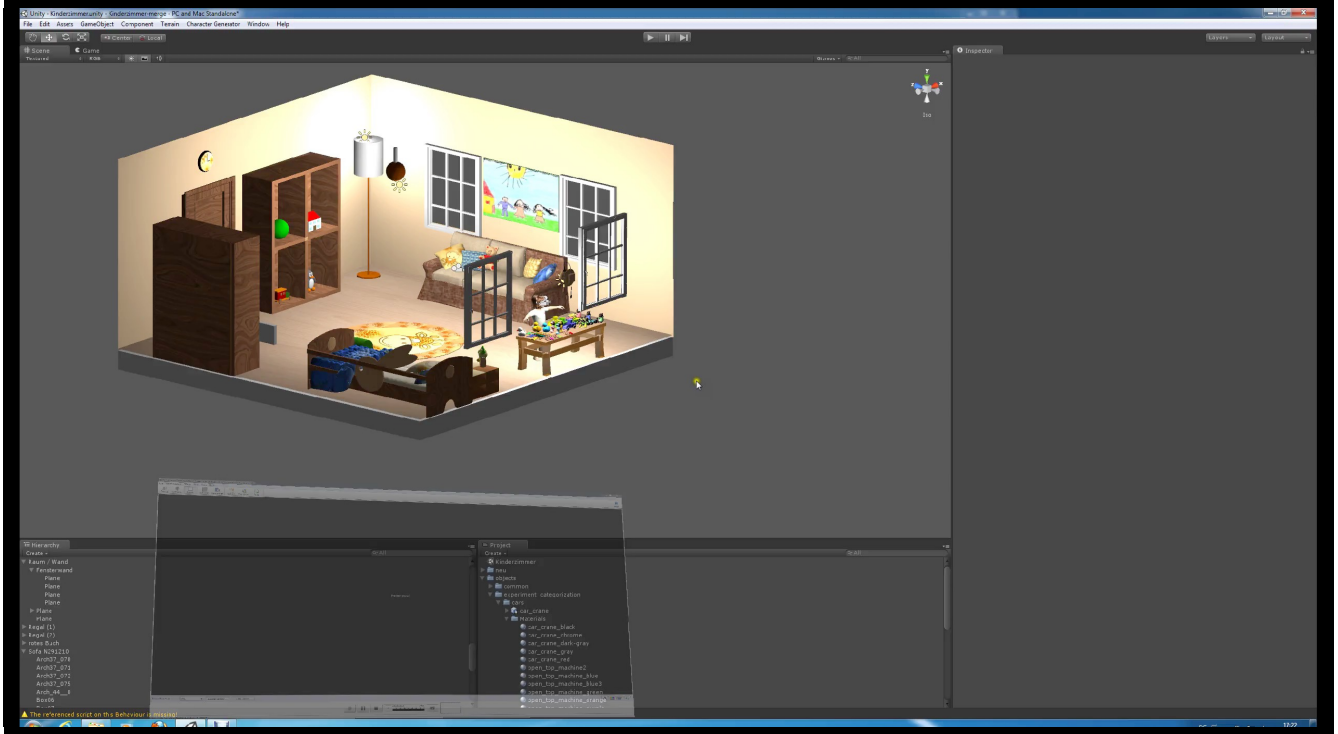
Neuro-informatics

- Neurosimulator ANNarchy
- BOLD fMRI signals

Machine learning

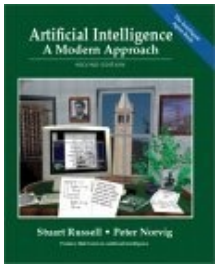
- Geometric deep learning
- Reservoir computing
- Cybersecurity and anomaly detection
- Reinforcement learning
- Facial expression recognition
- Natural language processing

Cooperation with psychologists, neuroscientists and physicians.




iit
ISTITUTO ITALIANO
DI TECNOLOGIA

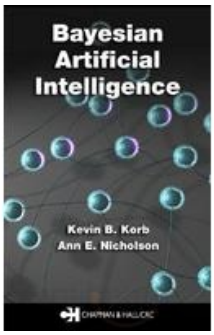
Literature



Stuart Russell and Peter Norvig
Artificial Intelligence: A Modern Approach
Prentice Hall, 2002 (Second Edition)

Künstliche Intelligenz: Ein moderner Ansatz.
München: Verlag Pearson Studium, 4., aktualisierte Auflage, 2023.

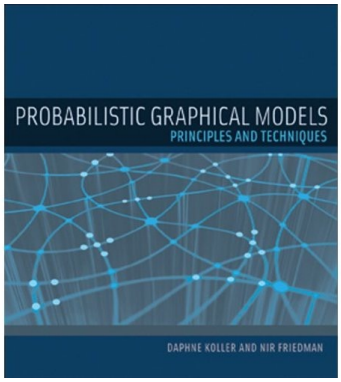
Wolfgang Ertel
Grundkurs Künstliche Intelligenz
Vieweg Verlag, 2008 (2. Auflage).



Kevin B Korb, Ann Nicholson,
Bayesian Artificial Intelligence
CRC Press, 2004.

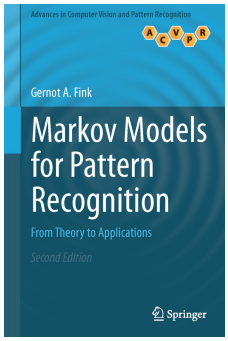
Literature

Bert-Uwe Köhler
Konzepte der statistischen Signalverarbeitung
Springer 2004

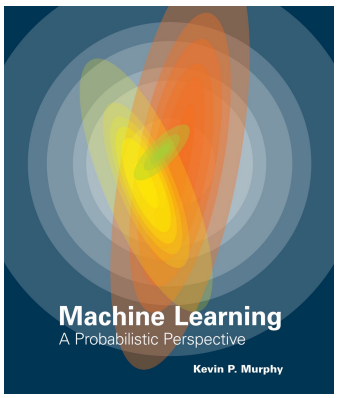


Daphne Koller & Nir Friedman
Probabilistic Graphical Models: Principles and Techniques
MIT Press 2009

Literature

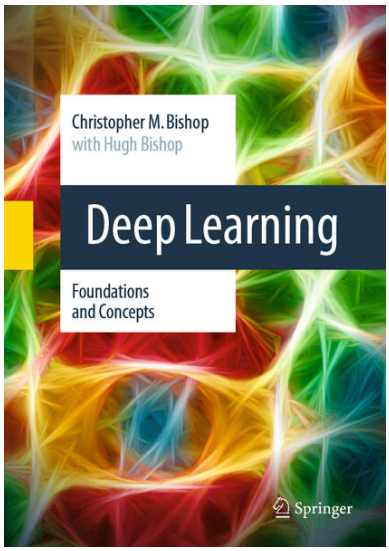


Gernot A. Fink
Markov Models for Pattern Recognition
Springer 2014

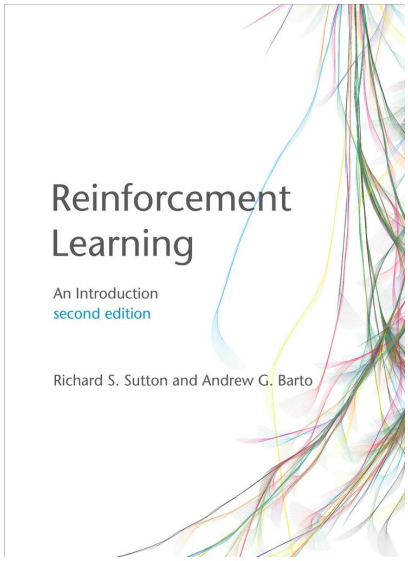


Kevin P. Murphy
Machine Learning: A Probabilistic Perspective
MIT Press 2012

Literature



Christopher M. Bishop, Hugh Bishop
Deep Learning
Foundations and Concepts
2023



Andrew G. Barto, Richard S. Sutton
Reinforcement Learning,
An Introduction
second edition, 2015