



Professur
Künstliche Intelligenz

Einführung in die künstliche Intelligenz I



CHEMNITZ UNIVERSITY
OF TECHNOLOGY

Prof. Dr. Fred Hamker, KI
Department of Computer Science

Einleitung 1

Einführung in die künstliche Intelligenz

Folien auf unserer Web-Seite
Additional set of slides in English on our web-page

Schriftliche Klausur in deutsch

und

Written exam will be also in english

Übungen: Dr. Michael Teichmann, Iliana Koulouri

Einleitung



Inhalt:

- Was ist Künstliche Intelligenz (KI)?
- Ziele der KI
- Geschichte der Künstlichen Intelligenz
- Welche Disziplinen beeinflussen die KI?
- Wo stehen wir?
- Gefahren der KI
- KI an der TU Chemnitz
- Literatur



Was ist Künstliche Intelligenz?

- Was ist Intelligenz?
- Wie kann man Intelligenz messen?
- Natürliche Intelligenz: Wie funktioniert unser Gehirn?

- Wie kann ich eine Maschine bauen, die sich verhält wie ein Mensch?
- Wie sehr will ich das Gehirn verstehen, oder kann man Intelligenz auch ganz anders modellieren?

Was ist Künstliche Intelligenz?

Ziel der KI ist es, Maschinen zu entwickeln, die sich intelligent verhalten.

John McCarthy (1955)



John McCarthy (1955)

Was ist Künstliche Intelligenz?

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 unvorhergesehenen Situation zurechtfinden können. Das ist etwas anderes, weshalb man den Begriff "intelligent" für Maschinen besser vermeiden sollte.

Peter Rechenberg, Informatik Spektrum 33(1), 2010

Was ist Künstliche Intelligenz?

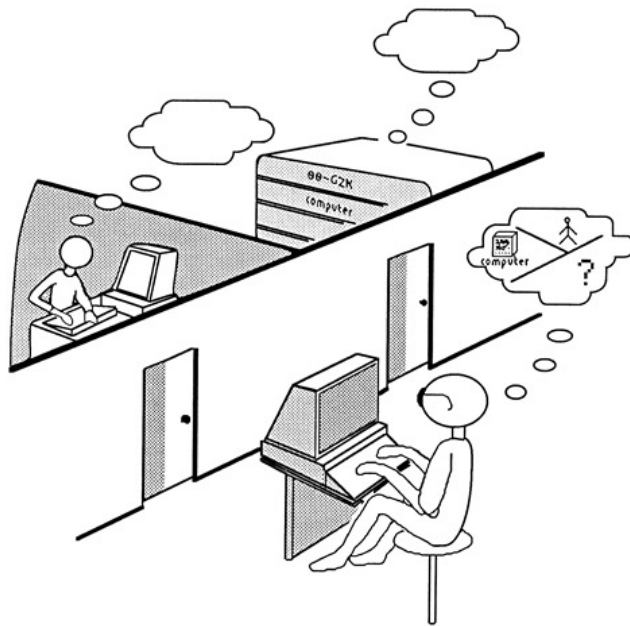
Starke KI:

Intelligenz erschaffen, die wie der Mensch nachdenken und Probleme lösen kann und die sich durch eine Form von Bewusstsein beziehungsweise Selbstbewusstsein sowie Emotionen auszeichnet.

Schwache KI:

Anwendungen von Interesse, zu deren Lösung nach allgemeinem Verständnis eine Form von "Intelligenz" notwendig zu sein scheint. (Simulation intelligenten Verhaltens mit Mitteln der Mathematik und der Informatik; es geht nicht um Schaffung von Bewusstsein oder um ein tieferes Verständnis von Intelligenz)

Der Turing-Test



Die Testperson Alice sitzt in einem abgeschlossenen Raum mit zwei Computerterminals. Ein Terminal ist mit der Maschine verbunden, das andere mit der Person Bob.

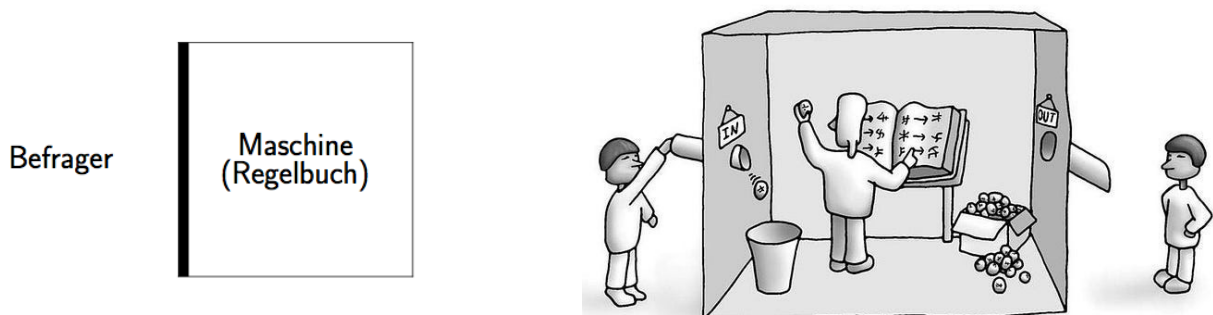
Alice kann Fragen stellen. Sie hat die Aufgabe, nach 5 Minuten zu entscheiden, an welchem Terminal die Maschine antwortet. Die Maschine besteht den Test, wenn sie Alice in 30% der Fälle täuschen kann.



Alan Turing (1912-1954)

Macht dieser Test heutzutage noch Sinn?
Alternative? Konkrete kognitive Aufgaben lösen.

Kritik von Searle am Turing-Test: Chinese Room

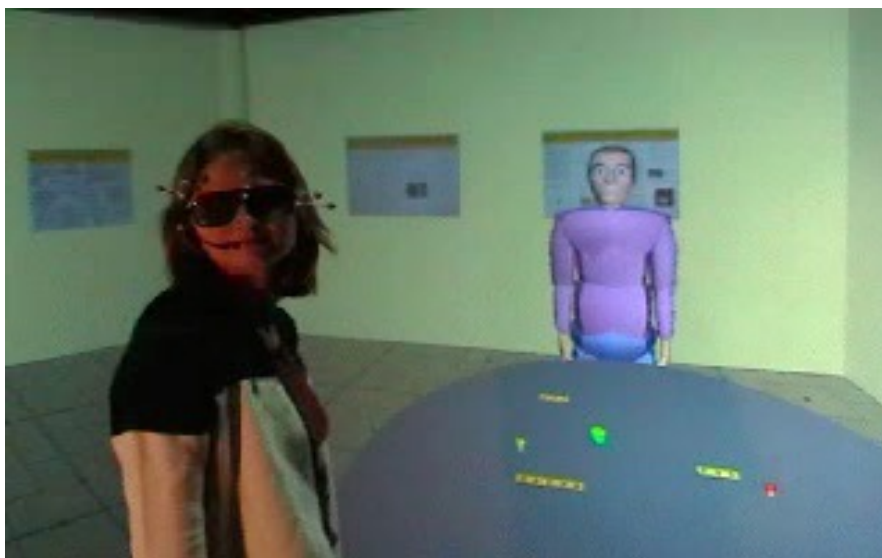


Stellen Sie sich vor, dass Sie kein Chinesisch sprechen, aber über ein riesiges Regelbuch verfügen, das es Ihnen ermöglicht, chinesische Sätze nachzuschlagen und nachher Ihnen sagt, wie sie auf Chinesisch antworten. Sie verstehen kein Chinesisch, können sich aber auf eine scheinbar intelligente Art verhalten.

Searle argumentierte, dass Computer, auch wenn sie intelligent erscheinen würden, es nicht wirklich wären, da sie nur so etwas wie das Regelbuch des chinesischen Zimmers benutzen würden. Sie verstehen nicht wirklich die Bedeutung des Inhalts.

Beispiel Affective Computing

Es gibt bisher keine Maschine, die wirkliche Emotionen hat, aber dennoch wird uns Menschen vorgetäuscht, als hätte diese Maschine Emotionen.



Der Turing-Test

Der Turing-Test definiert folgende Arbeitsdisziplinen innerhalb der KI:

1. Wissensrepräsentation: Speicherung von evtl. komplex strukturierten Informationen
2. Schließen und logisches Denken: Beantwortung von Fragen und Inferenz unter Verwendung der gespeicherten Informationen
3. Verarbeitung geschriebener natürlicher Sprache: Verstehen, Generieren, Übersetzen geschriebener Sprache
4. Maschinelles Lernen: Anpassung an neue Situationen, Erkennung von Mustern und Extrapolation

Der klassische Turing-Test vermied bewusst direkte physische Interaktion zwischen Befrager und Computer (physische Simulation einer Person nicht erforderlich für die Intelligenz).

Totaler Turing-Test: Begriff Intelligenz vollständiger erfassen

5. Computer Vision: visuelle Wahrnehmung der Umwelt
6. Verstehen und Synthese gesprochener Sprache
7. Robotik: Handling von Objekten, Fortbewegung

Diese sieben Disziplinen bilden den größten Teil der KI

Ziele der KI

Langfristiges Ziel:

Systeme zu bauen, welche die Leistungsfähigkeit des Menschen auf breiter Basis erreichen oder übertreffen. Ob dieses Ziel grundsätzlich erreichbar (und ethisch/wirtschaftlich erwünscht) ist, wird seit langem heftig diskutiert.

Kurzfristiges Ziel:

Spezialisierte Systeme zu bauen, welche ganz bestimmte Aufgaben lösen können.

Geschichte der KI

- Frühe Einflüsse
 - Gödel (1931) Vollständigkeit der Prädikantenlogik erster Stufe
- Reifungsprozess der KI (1943 – 1955)
 - Modell eines künstlichen neuronalen Netzes (McCulloch & Pitts, 1943)
 - Hebb'sche Lernregel von Neuronen (Hebb 1949)
 - Alan Turing (1950)
 - Minsky (1951) Neuronenrechner: 40 Neuronen mit 3000 Röhren
- Geburt der KI als eigene Disziplin, Workshop am Dartmouth College (1956)
 - McCarthy: LISP, Programmiersprache, die speziell für die Verarbeitung von symbolischen Strukturen erschaffen wurde
 - Newell & Simon: Logic Theorist: erster automatischer Theorembeweiser; Zeigten dass Computer, die sonst mit Zahlen rechnen, auch Symbole verarbeiten können.

Geschichte der KI

- Früher Enthusiasmus, große Erwartungen (1952 – 1969)
 - Newell & Simon (1961) General Problem Solver (GPS) – A program that simulates human thought
 - Samuel (1959) Dameprogramm
 - Robinson (1965) Resolutionsmethode (Widerspruchsbeweis)
 - Weizenbaum (1966) Eliza: Antwortet wie ein menschlicher Psychologe
 - Blockwelt
 - Perceptron
 - Suchverfahren
 - Erwartungen wurden stark hochgespielt
 - wenig oder gar kein Wissen über den Anwendungsbereich

Geschichte der KI

- Konfrontation mit der Realität (1966 – 1973)
 - Systeme scheiterten bei schwierigeren Aufgaben (maschinelle Übersetzung)
 - Komplexität vieler Probleme
 - Minsky & Papert (1969) Perceptron kann das ‚xor‘ Problem nicht lösen (-> Eiszeit der neuronalen Netze)
- Wissensbasierte, Expertensysteme (1969 – 1979)
 - de Dombal (1972) Expertensystem zur Diagnose von Bauchkrankheiten
 - Colmerauer (1972) Logikprogrammiersprache PROLOG
 - Shortliffe & Buchanan (1976) MYCIN: medizinisches Expertensystem (ca. 450 Regeln; kann mit einigen Experten gleichziehen; besser als Jungmediziner)

Geschichte der KI

- KI wird zu einem Industriezweig (1980 bis heute)
 - Erste kommerzielle Expertensysteme.
 - Automatisierte logistische Planung und Zeitpläne für Transportaufgaben.
 - Japan startet mit großem Aufwand das „5th generation project“ mit dem Ziel Leistungsfähige PROLOG-Maschinen zu bauen.
 - Big Data (Google, Facebook, Microsoft)
 - Autonomes Fahren
 - Large Language Modelle (LLM)

- Die Rückkehr der neuronalen Netze (1986 bis 2000)
 - Enttäuschung über die Möglichkeiten der Expertensysteme
 - Rumelhart, Hinton & Williams (1986) Backpropagation
 - Kohonen (1995) Self-Organizing Feature Maps (SOM)
 - Fukushima (1980) Neokognitron, NN zur Objekterkennung
 - Convolutional Networks (LeCun, 1990)

Geschichte der KI

- Statistische Ansätze (1987 bis heute)
 - Hidden Markov Modelle
 - Pearl (1988) Probabilistische Graphische (Bayes) Netze
 - Vapnik (1995) Support Vektor Maschine – Klassifikator auf Basis einer statistischen Lerntheorie

- Reinforcement Lernen (1990 bis heute)
 - Lernen von Agenten, die mit der Umwelt in Interaktion stehen, über unspezifisches Feedback (Barto & Sutton, 1998)

- Computational Neuroscience (1995 bis heute)
 - Enttäuschung über die Möglichkeiten klassischer neuronaler Netze, insbesondere hinsichtlich der Erklärung menschlicher Fähigkeiten
 - Aufkommen von Methoden, Neuronenverbände nach biologischem Vorbild zu simulieren
 - derzeit eher biologisch ausgerichtet (EU: Human Brain Project), aber erste Ansätze als Technologie

Geschichte der KI

- Das Entstehen intelligenter Agenten (1995 bis heute)
 - Roboter, Sensoren, Modellierung von Wahrscheinlichkeiten
 - Roboterfußball
 - Urban Challenge (Darpa USA)
 - Autonomes Fahren

Geschichte der KI

- Deep Learning Architekturen (2010 bis heute)
 - Deep Learning Netzwerke gewinnen alle Wettbewerbe in Mustererkennung (IJCNN 2011 Traffic Sign Recognition Competition, ICPR 2012 Mitosis Detection in Breast Cancer Histological Images, ImageNet 2012 Large Scale Visual Recognition Challenge).
 - In 2014 kauft Google für 400 M€ Deepmind.
 - Deep RL Learning, vom Bild zur Robotersteuerung in einer Architektur
 - GPT, ChatGPT: GPT (**G**enerative **P**re-trained **T**ransformer) sind Sprachmodelle, die statistische Abhängigkeiten von Wörtern oder Satzbausteinen benutzen und so eine komplexe Kommunikation ermöglichen. Sie erlauben eine sehr komplexe Kommunikation und das Verfassen von Texten, Programmcode und Bildern. Transformer sind eine bestimmte Architektur der Neuronalen Netze.

Beispiel: Deep RL Learning, vom Bild zur Robotersteuerung in einer Architektur



Ein Schachcomputer gewinnt gegen den Schachweltmeister Kasparov – Mai 1997



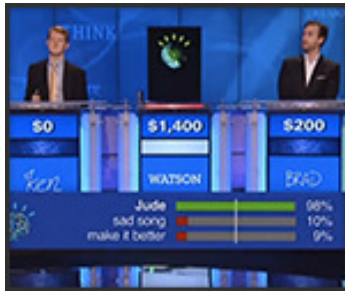
1:0
0:1
1/2:1/2
1/2:1/2
1/2:1/2
0:1



Damit war das System der erste Computer, der einen Wettkampf unter Turnierbedingungen gegen einen amtierenden Schachweltmeister für sich entscheiden konnte.

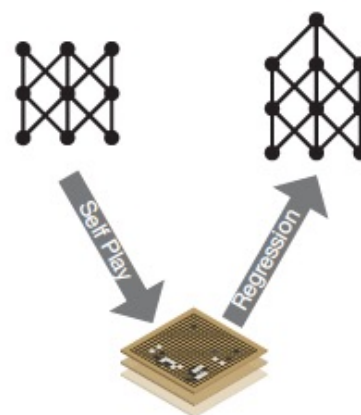
- Spielstärke aus enormer Rechenleistung: 100 – 200 Mio Stellungen pro Sek.
- Stellungsbewertung

IBM-Supercomputer Watson gewinnt Quizshow Jeopardy – Februar 2011



- Aufgabe: passende Frage zu einem – oft kryptischen – Hinweis (meist bewusst mehrdeutig formuliert) formulieren.
 - Verknüpfung mehrerer Fakten erforderlich
 - Punkten kann nur, wer schneller ist als seine beiden Konkurrenten.
 - im Vergleich zu Jeopardy (<http://www.jeopardy.com/>) ist Schach ein mathematisch leicht zu definierendes Spiel
- Sucht im Archiv nach Wörtern, die mit der Anfrage (als Text) in Verbindung stehen
 - Wählt 50 bis 60 Informationseinheiten aus und erstellt aus maximal 200 Hypothesen ein Ranking
 - **2015: IBM Watson Health geht an den Start**
 - **Januar 2022, IBM scheitert und verkauft Watson Health**

AlphaGo gewinnt gegen den Go „Weltmeister“ Lee Sedol – März 2016



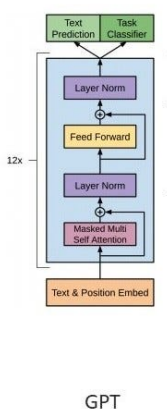
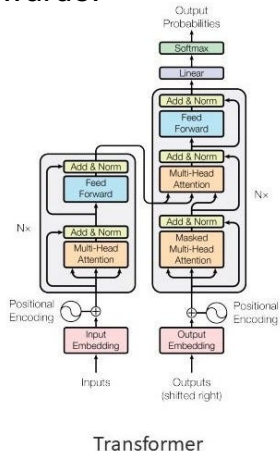
Go galt aufgrund seiner hohen Anzahl an möglichen Spielzügen und damit einer hohen Verzweigungsrate in Suchbäumen als nicht für die KI gewinnbar. Weiterhin ist die Bewertung eines Spielzustandes deutlich schwieriger als beim Schach, wo allein die Figuren an sich schon einen Wert haben.

AlphaGo gewinnt gegen den Go „Weltmeister“ Lee Sedol – März 2016

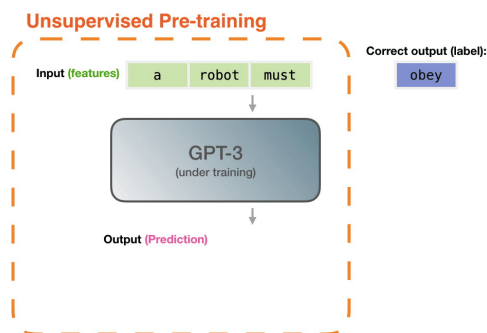
- Kombiniert Monte Carlo Tree Search (MCTS) mit lernfähigen Neuronalen Netzen und Reinforcement Learning
- MCTS: Methode zum Aufbau eines gewichteten Suchbaums durch viele Zufallsspiele (schnell)
- Policy Network: Bestimmt die Aktion
- Value Network: Schätzt den Wert eines Zustandes
- Überwachtes Lernen von policy networks (Experten)
- Reinforcement Learning von policy networks (Self-play)
- Reinforcement Learning von value networks
- Kombination von MCTS, Policy und Value Netzwerk
- Hardware (Version in Nature: 1202 CPUs und 176 GPUs)

ChatGPT

GPT (Generative Pre-trained Transformer) ist ein Sprachmodell (neuronales Netz) welches von OpenAI auf Basis von Transformer Neuronalen Netzen entwickelt wurde.



„Human-like conversation“, welches durch „next-word prediction“ trainiert wird.

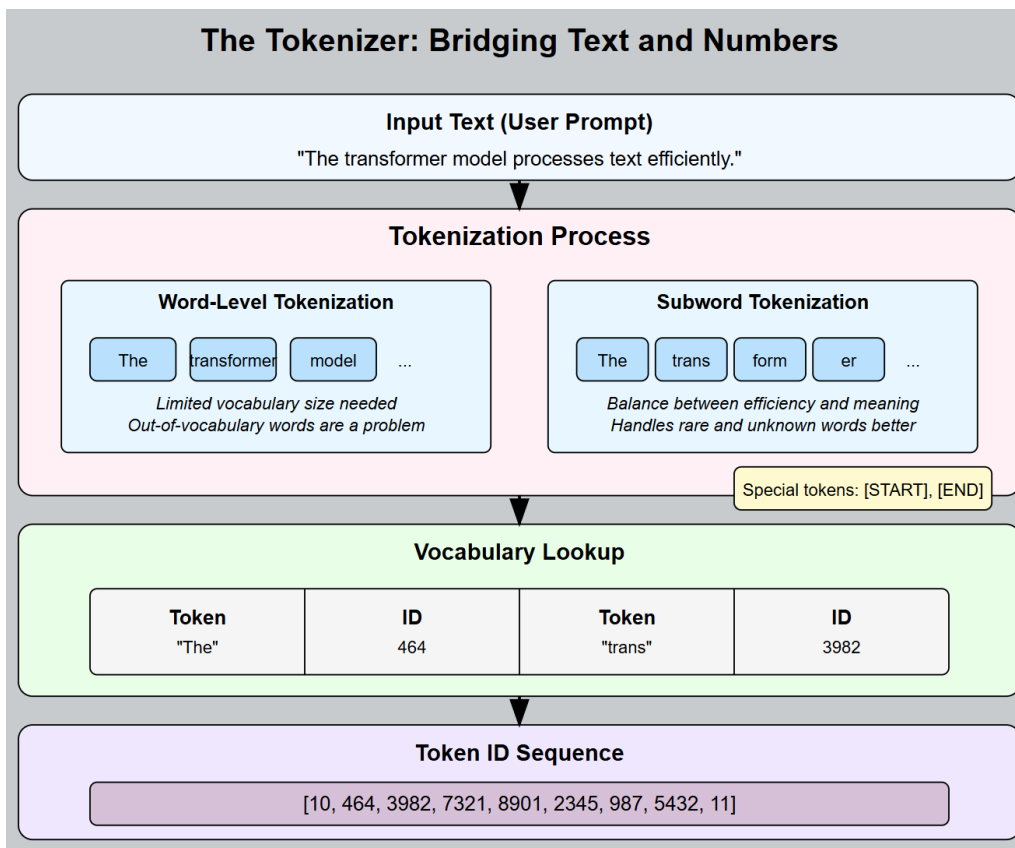


Revolution or just a bullshit Parrot ?

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

Chat GPT wurde mit Hilfe von supervised und reinforcement learning Methoden verbessert

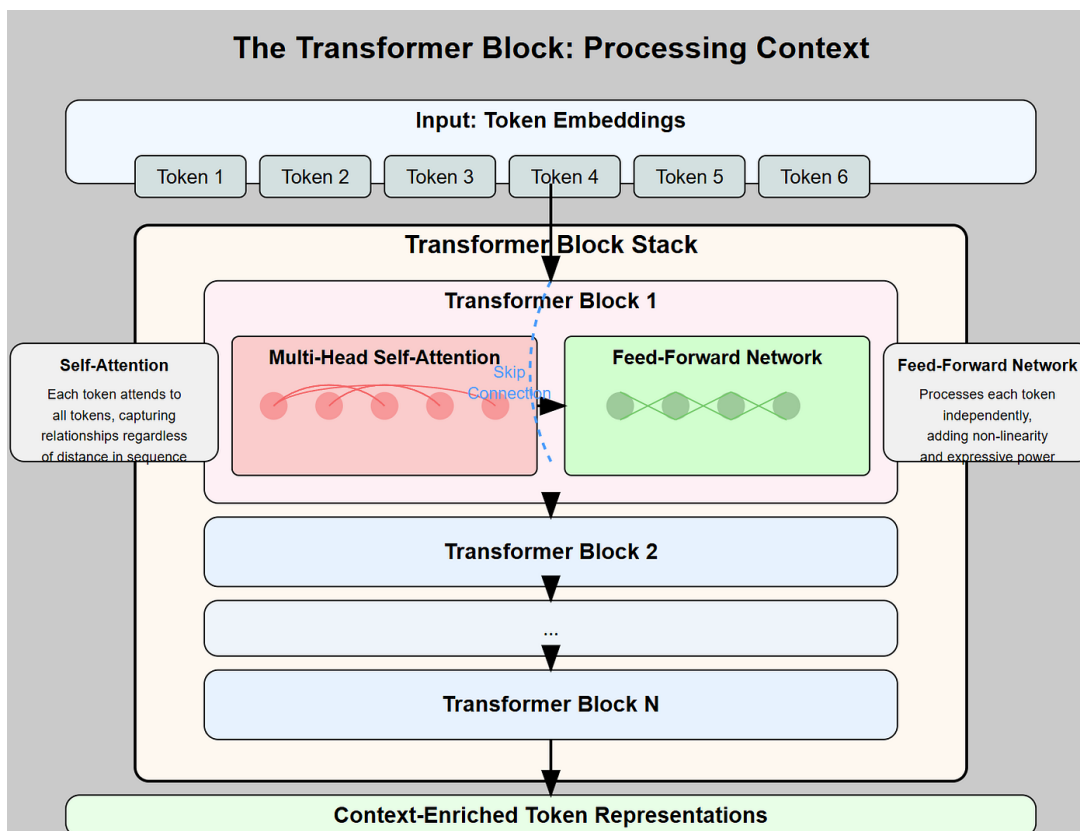
Large Language Modelle



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

Large Language Modelle

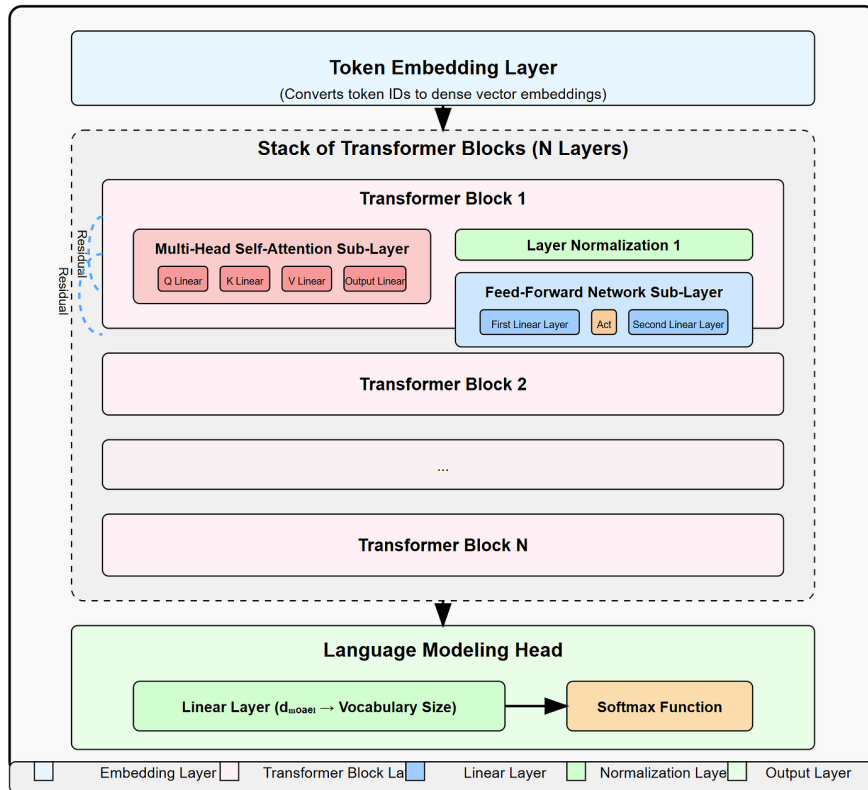
knowledge is distributed and emergent across the entire set of parameters



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

Large Language Modelle

Layers in Transformer LLM Architecture

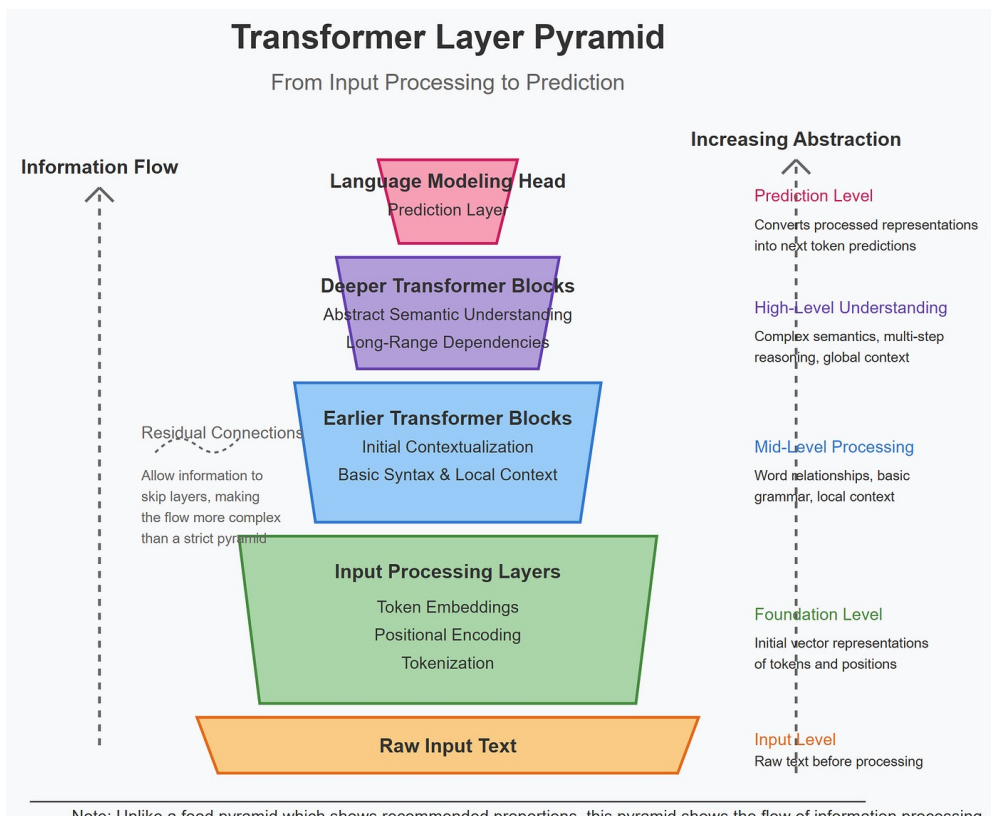


<https://medium.com/codex/how-transformer-llms-generate-text-one-token-at-a-time-5531838bc2a1>

Large Language Modelle

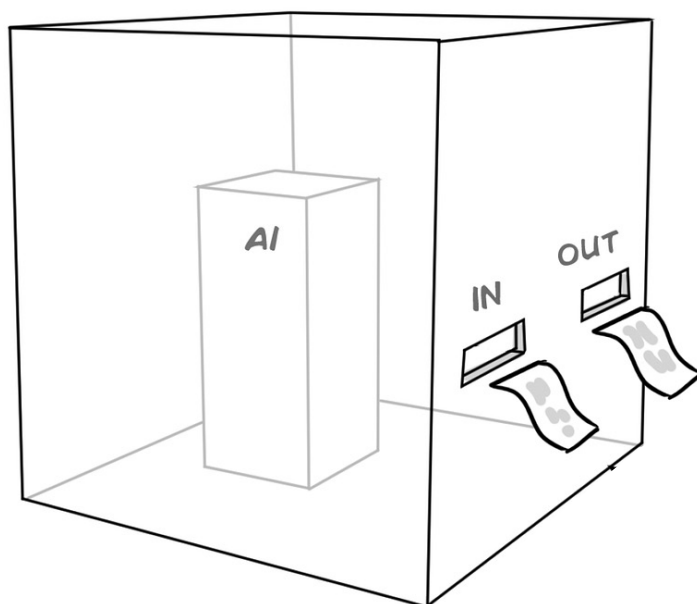
Transformer Layer Pyramid

From Input Processing to Prediction



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

Large Language Modelle und Chinese Room



Large Language Modelle kommunizieren alleinig über Text (Tokens). Im Grunde werden sie darauf trainiert, nächste oder maskierte Wörter vorherzusagen.

Sind sie intelligent?

Welche Disziplinen beeinflussen die KI ?

- Philosophie (428 v. Chr. bis heute)
- Mathematik (ca. 800 bis heute)
 - Prädikatenlogik
 - Algorithmus
 - Wahrscheinlichkeitstheorie
 - Spieltheorie
 - Markov'sche Entscheidungsprozesse
- Neurowissenschaft (1861 bis heute)
 - Wie verarbeitet das Gehirn Informationen?
 - Einzelzelleableitungen, bildgebene Verfahren
 - Computational Neuroscience

Welche Disziplinen beeinflussen die KI ?

- Künstliche Neuronale Netze (1941 bis heute)
 - Backpropagation Lernalgorithmus
 - Deep Learning Architekturen
 - Transformer Neuronale Netze (Encoder, Decoder, Self-attention)
- Psychologie (1879 bis heute)
 - Wie denken und handeln Menschen und Tiere?
 - Kognition
 - Patientenstudien
- Technische Informatik (1940 bis heute)
 - künstliche Maschinen (Computer, Roboter)
 - Neuromorphic Computing
- Kontrolltheorie und Kybernetik (1948 bis heute)
- Linguistik (1957 bis 2010)

Inwieweit haben wir das Ziel, intelligente Maschinen bauen zu können erreicht?

- Ein Computer kann spezielle Aufgaben besser bearbeiten als der Mensch.
- Computer werden immer schneller und damit können auch schwierige Aufgabenstellungen durch Probieren und Simulieren gelöst werden.
- Heutige Computer sind allerdings mit der Interpretation einer Szene und dem Fällen von Entscheidungen in der realen Welt (vollkommen) überfordert, werden aber immer besser.
- Menschen können viel besser lernen als Computer
- Neuronale Netze können in Form von Sprachmodellen eine riesige Wissensbasis haben und komplexe Texte verfassen sowie Bilder und Videos generieren.

Inwieweit wollen wir Künstliche Intelligenz

- Menschen könnten auf Grund der zunehmenden Automatisierung ihre Arbeit verlieren.
- Menschen könnten das Selbstverständnis verlieren, einzigartig zu sein.
- Menschen könnten einen Teil ihrer Persönlichkeitsrechte verlieren.
- Die Verwendung von Systemen aus KI könnte zu einem Verlust der Verantwortung führen (auton. Fahrzeuge, Flugzeuge (Boeing 737 Max 8)).
- Erste Tote durch autonome Fahrzeuge (Tesla, 2016, 2018, Uber, 2018)

- Der Erfolg der KI könnte das Ende der menschlichen Rasse bedeuten. Fast jede Technologie hat das Potential, in den falschen Händen Schaden anzurichten (KI gesteuerte Waffen).

Inwieweit wollen wir Künstliche Intelligenz

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 hÉigartaigh^[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]

<p>¹ Corresponding author. miles.brundage@philosophy.ox.ac.uk Future of Humanity Institute, University of Oxford; Arizona State University</p> <p>² Corresponding author. s4472@msu.ac.uk Centre for the Study of Existential Risk, University of Cambridge</p> <p>³ OpenAI</p> <p>⁴ Open Philanthropy Project</p> <p>⁵ Electronic Frontier Foundation</p> <p>⁶ Future of Humanity Institute, University of Oxford</p> <p>⁷ Future of Humanity Institute, University of Oxford; Yale University</p> <p>⁸ Center for a New American Security</p>	<p>⁹ American University</p> <p>¹⁰ Endgame</p> <p>¹¹ Endgame</p> <p>¹² University of Oxford/Arizona State University/New America Foundation</p> <p>¹³ Center for a New American Security</p> <p>¹⁴ Stanford University</p> <p>¹⁵ Future of Humanity Institute, University of Oxford</p> <p>¹⁶ Centre for the Study of Existential Risk and Centre for the Future of Intelligence, University of Cambridge</p> <p>¹⁷ Centre for the Study of Existential Risk, University of Cambridge</p>	<p>¹⁸ Centre for the Study of Existential Risk, University of Cambridge</p> <p>¹⁹ Future of Humanity Institute, University of Oxford</p> <p>²⁰ Future of Humanity Institute, University of Oxford</p> <p>²¹ Information Society Project, Yale University</p> <p>²² Future of Humanity Institute, University of Oxford</p> <p>²³ OpenAI</p> <p>²⁴ University of Bath</p> <p>²⁵ University of Louisville</p> <p>²⁶ OpenAI</p>
---	--	--

Aktive Forscher der KI Szene machten sich Gedanken über die nahe Entwicklung und potentiellen Gefahren der KI durch Missbrauch.

Zu ihren Lösungsvorschlägen gehören auch die Zentralisierung und nicht Verbreitung der Technologie, solange Gefahren zu hoch sind.

Siehe auch EU KI-Gesetz:
<https://www.europarl.europa.eu/topics/de/article/20230601/STO93804/ki-gesetz-erste-regulierung-der-kunstlichen-intelligenz>

Inwieweit wollen wir Künstliche Intelligenz

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

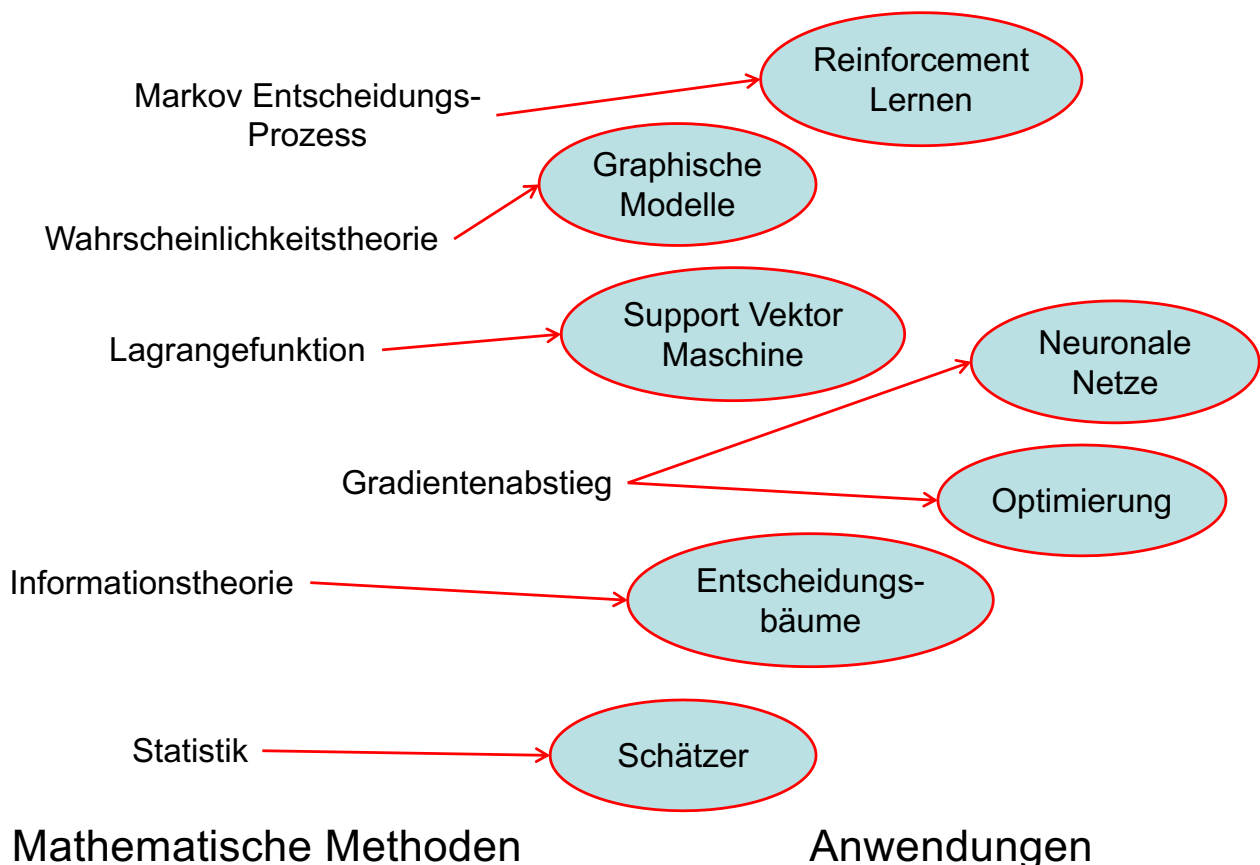
„Contemporary AI systems are now becoming human-competitive at general tasks, ...: Should we let machines flood our information channels with propaganda and untruth? Should we automate away all the jobs, including the fulfilling ones? Should we develop nonhuman minds that might eventually outnumber, 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.“

Eine Pause gab es vermutlich nicht wirklich ...

Übersicht – Einführung in die KI

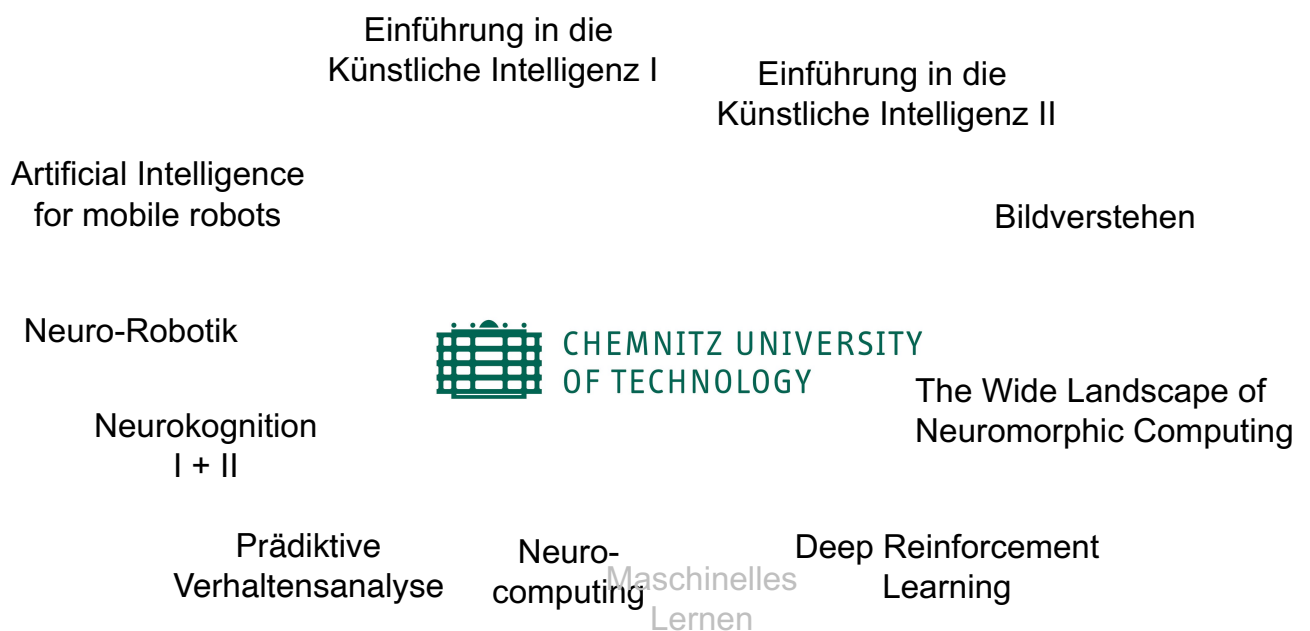


Übersicht – Einführung in die KI I

Inhalt:

- Begriffe
- Suchen, Spielen, Probleme Lösen
- Wahrscheinlichkeitstheorie
- Optimierung
- Neuronale Netze
- Support Vektor Maschine
- Reinforcement Lernen
- Statistik und Schätzer
- Informationstheorie
- Entscheidungsbäume

KI an der Informatik der TU Chemnitz



Siehe auch Studiengang Master Neurorobotik

Forschung an der Professur KI

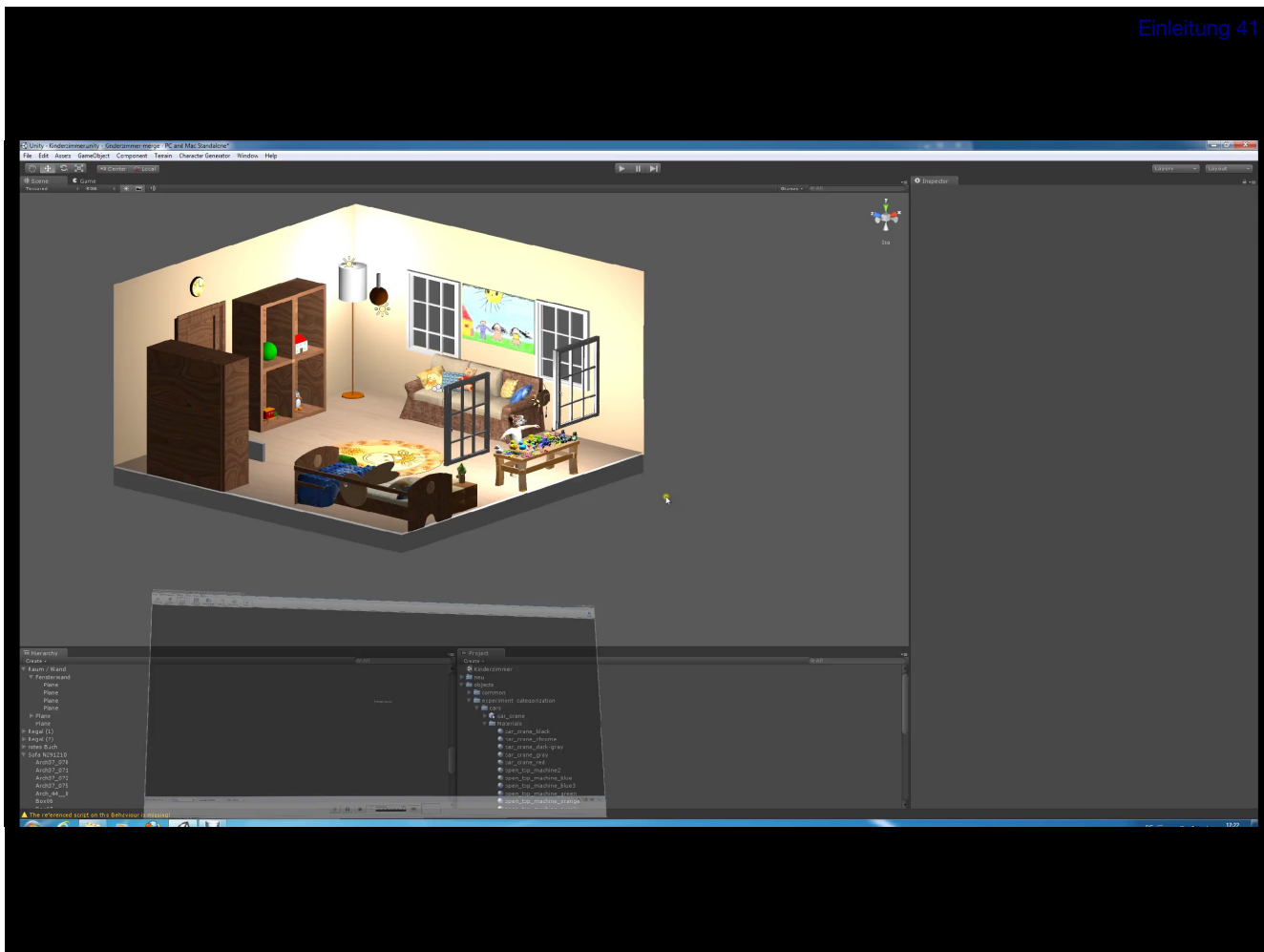
Forschung:

- Computational neuroscience
 - Perception
 - Cognition and action
- Neuro-cognitive agents
- Machine learning

Entwicklung:

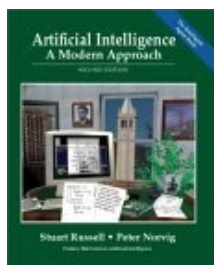
- ANNarchy (Neuro-Sim)
- Parallele Programmierung (Multi-Core, GPU, Neuromorph)
- Agenten in VR (Unity)
- Robotik (BodyOwnership)

Kooperationen mit Maschinenbau, Psychologen, Neurowissenschaftlern und Medizinern.





Literatur



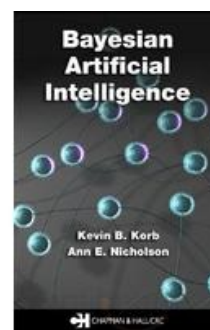
Stuart Russell and Peter Norvig
 Artificial Intelligence: A Modern Approach
 Prentice Hall, 2002 (Second Edition)

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



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

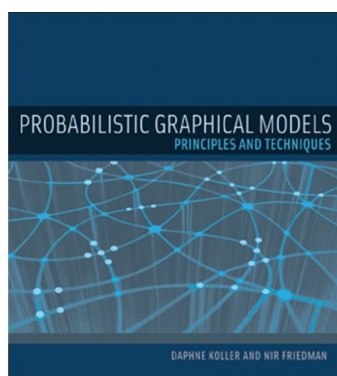
Kevin B Korb, Ann Nicholson,
 Bayesian Artificial Intelligence
 CRC Press, 2004.



Literatur

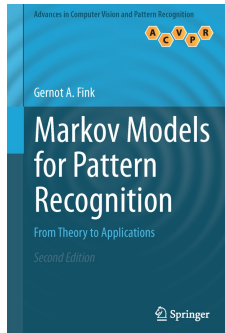


Bert-Uwe Köhler
 Konzepte der statistischen Signalverarbeitung
 Springer 2004

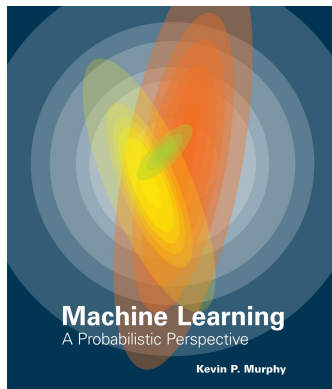


Daphne Koller & Nir Friedman
 Probabilistic Graphical Models: Principles and Techniques
 MIT Press 2009

Literatur

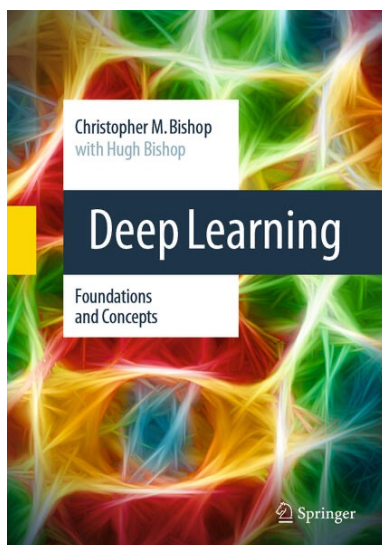


Gernot A. Fink
Markov Models for Pattern Recognition
Springer 2014

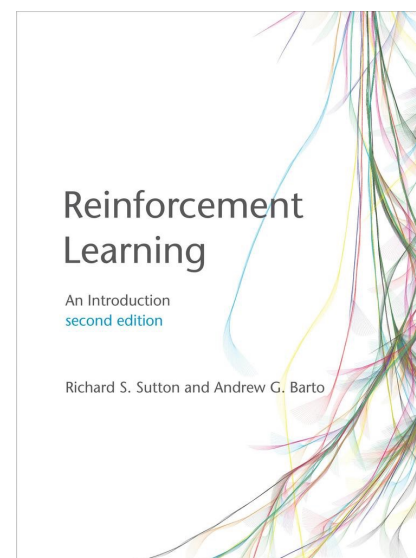


Kevin P. Murphy
Machine Learning: A Probabilistic Perspective
MIT Press 2012

Literatur



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



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