

Künstliche Intelligenz

Techniken, Chancen, Risiken

Dr. Manuela Lenzen

KI goes Business, Fachhochschule Wiener Neustadt, 02. Juni 2023

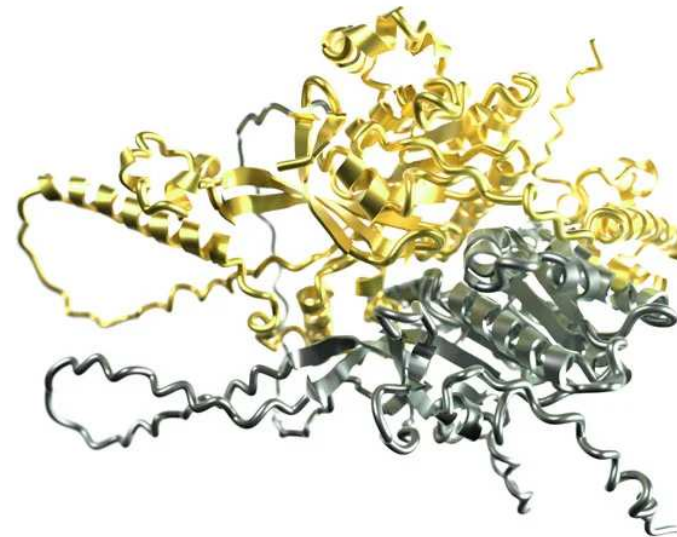
Verheißungen: Wissenschaft

2021 BREAKTHROUGH OF THE YEAR

Protein structures for all

AI-powered predictions show proteins finding their shapes

BY ROBERT SERVICE



https://www.science.org/content/article/breakthrough-2021#section_breakthrough

Verheißungen: Arbeitswelt

- Optimierung
- Effizienz
- Beschleunigung
- „Losgröße 1 zum Preis von Massenware“
- Wider den Arbeitskräftemangel
- Einsparung von Personalkosten

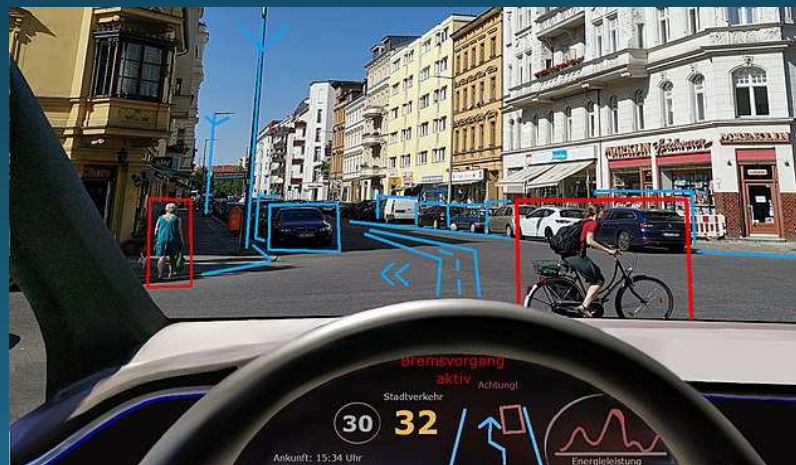
Verheißungen: Alltag



© Bretislav Valek, Wikimedia



© CITEC, Uni Bielefeld



© Eschenzweig,
Wikimedia

Verheißungen: Militär



Predator © Wikimedia



UGV: unmanned ground vehicle,
Milrem Robotics, © Wikimedia

Ängste (alt)

Experte erklärt: Ein Milliardenmarkt könnte sich zur größten Bedrohung unserer Zeit entwickeln

Christoph Damm

🕒 17.07.2017, 17:22 📈 12,544

INTELLIGENTE COMPUTER

So gefährlich wie Atomwaffen

VON JAY TUCK am 9. August 2016

"Das größte Risiko für unsere Zivilisation": Elon Musk warnt erneut vor KI

(Heise News, 17.07.2017)

TERMINATOR 3

REBELLION DER MASCHINEN

Stephen Hawking: "KI könnte das schlimmste Ereignis der Menschheit werden"

(Futurezone, 07.11.2017)

Ängste (neu)



„Es ist schwer vorstellbar, wie man Kriminelle davon abhalten können wird, KI für böse Dinge zu missbrauchen.“

(Geoffrey Hinton in der *New York Times*)

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

27565

Add your
signature

Verwirrung I: Science-Fiction

- Geschichten über künstliche Menschen und andere Automaten
- von der Antike, z.B.:
der Wächter des Hephaistos
- bis zur Science-Fiction, z.B.
Hal 9000, A Space Odyssey



Sagenhafter künstlicher Wächter
des Hephaistos ©Wikipedia

Verwirrung II: Anthropomorphismus



Sophia, Hanson Robotics

© ITU Pictures from Geneva, Switzerland –Wikimedia

Was ist Künstliche Intelligenz?

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.

John McCarthy, Marvin L. Minsky, Nathaniel Rochester, and Claude E. Shannon, AI Magazine, Vol. 27 , NO 4 (2006), orig.: 31. August 1955

I. Anwendungen

- Übersetzungsprogramme
- Programme zur Analyse von Bildern
- Sprachprogramme
- Vergabe von Werbeplätzen
- Verbesserung der Wetter- und Klimasimulation
- Entdecken von illegalen Transaktionen
- Frühes Erkennen von Herzinfarkten
- Prognose der Lebenserwartung
- Persönlichkeitsanalyse bei Bewerbungen
- Roboter für die Industrieproduktion
- Soziale Roboter
- ...

II. Die Natur verstehen



Stabheuschrecke
© Wikipedia



Hector, © Universität Bielefeld

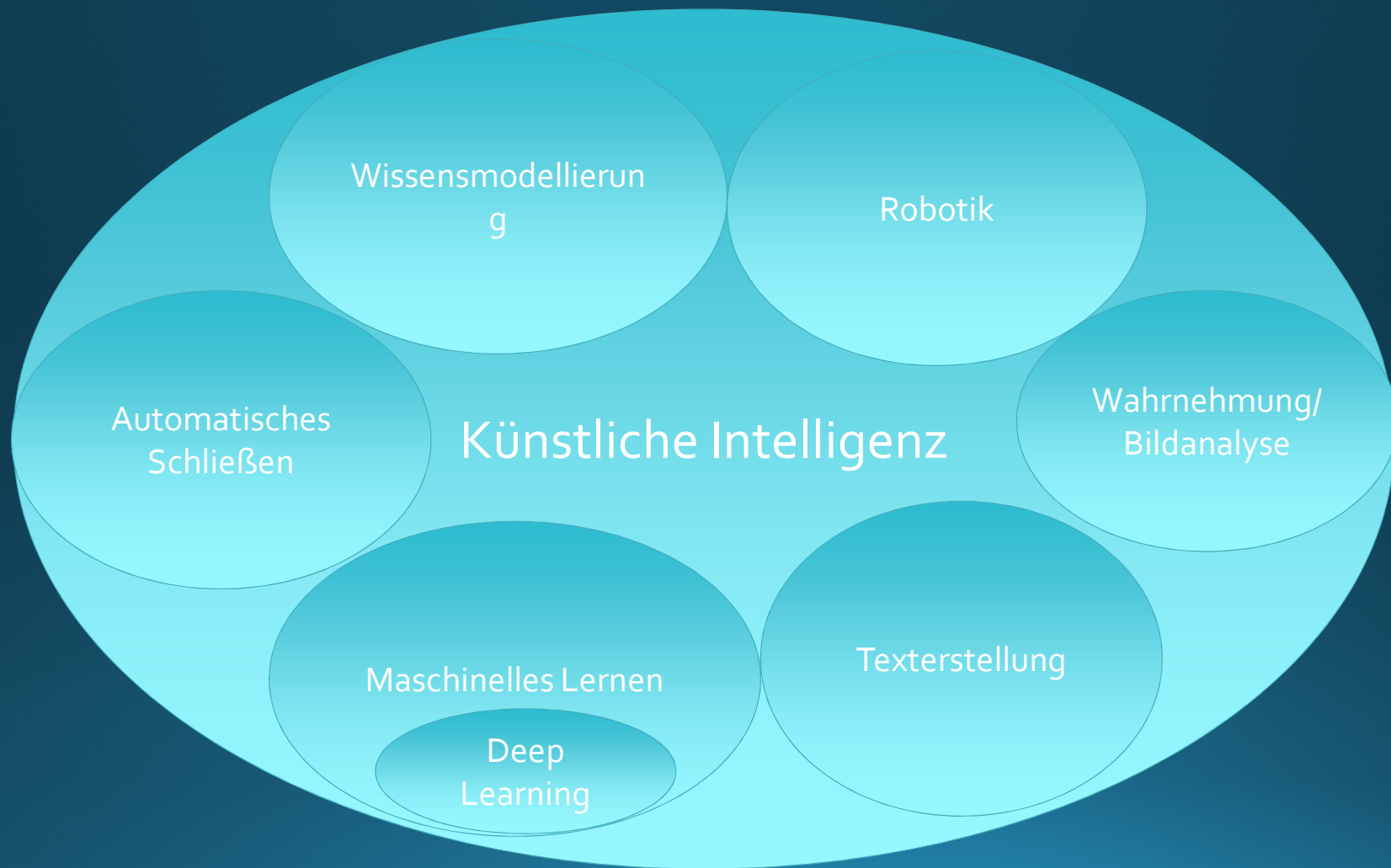
III. Künstliche Allgemeine Intelligenz Artificial General Intelligence (AGI) / Human Level Intelligence

- Orientierung in der Welt
- Handlungsplanung
- Nicht auf eine bestimmte Anwendung spezialisiert
- Lernfähig
- Ähnlich wie der Mensch oder besser

„Künstliche Intelligenz ist die Wissenschaft und Technik von der **Herstellung intelligenter Maschinen**, insbesondere intelligenter Computerprogramme. Sie ist verwandt mit der ähnlichen Aufgabe, Computer zu nutzen, um die **menschliche Intelligenz zu verstehen**, aber die KI muss sich nicht auf Methoden beschränken, die biologisch beobachtbar sind.“

John McCarthy, 2004

Forschungsfeld Künstliche Intelligenz



Wissensbasierte Systeme

MYCIN Sample Rule

Human-Readable Format

```
IF    the stain of the organism is gram negative
AND   the morphology of the organism is rod
AND   the aerobiocity of the organism is gram anaerobic
THEN  there is strong evidence (0.8)
      that the class of the organism is enterobacteriaceae
```

MYCIN Format

```
IF    (AND (SAME CTEXT GRAM GRAMNEG)
          (SAME CTEXT MORPH ROD)
          (SAME CTEXT AIR AEROBIC))
THEN  (CONCLUDE CTEXT CLASS ENTEROBACTERIACEAE
      TALLY .8)
```

[Durkin 94, p. 133]

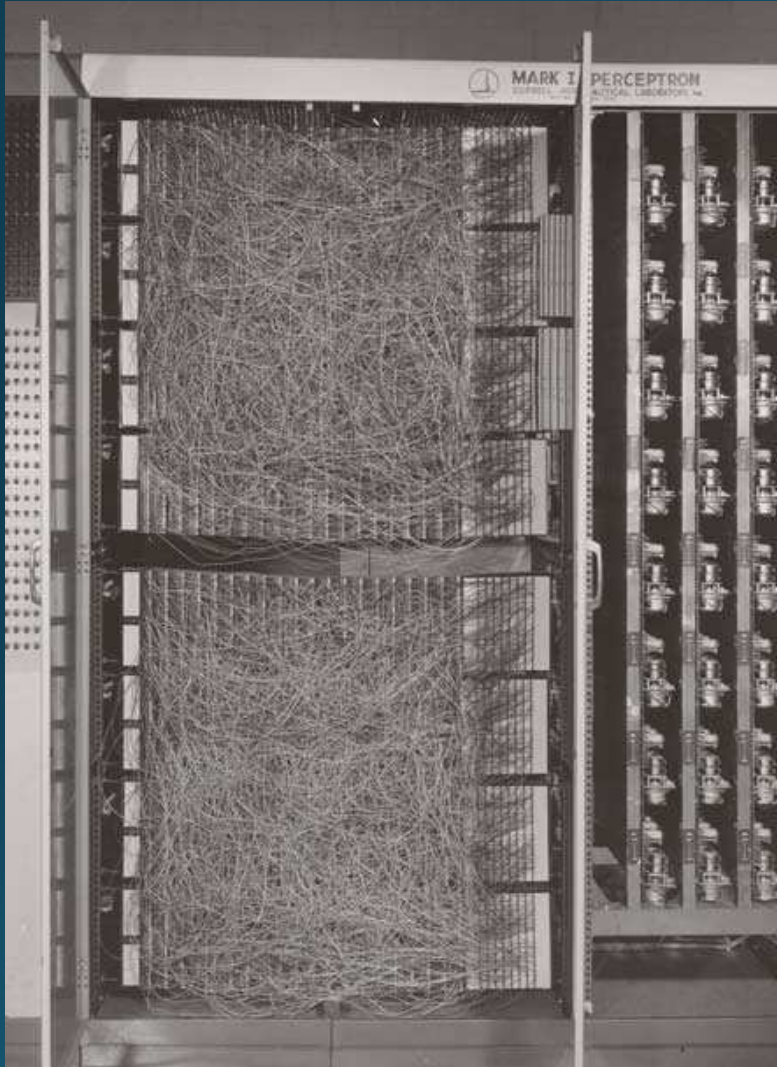
Good Old
Fashioned
Artificial
Intelligence:
GOFAI

© Durkin, J. (1994): Expert Systems Design and Development.
Prentice Hall, Englewood Cliffs, NJ (1972)

Lernverfahren



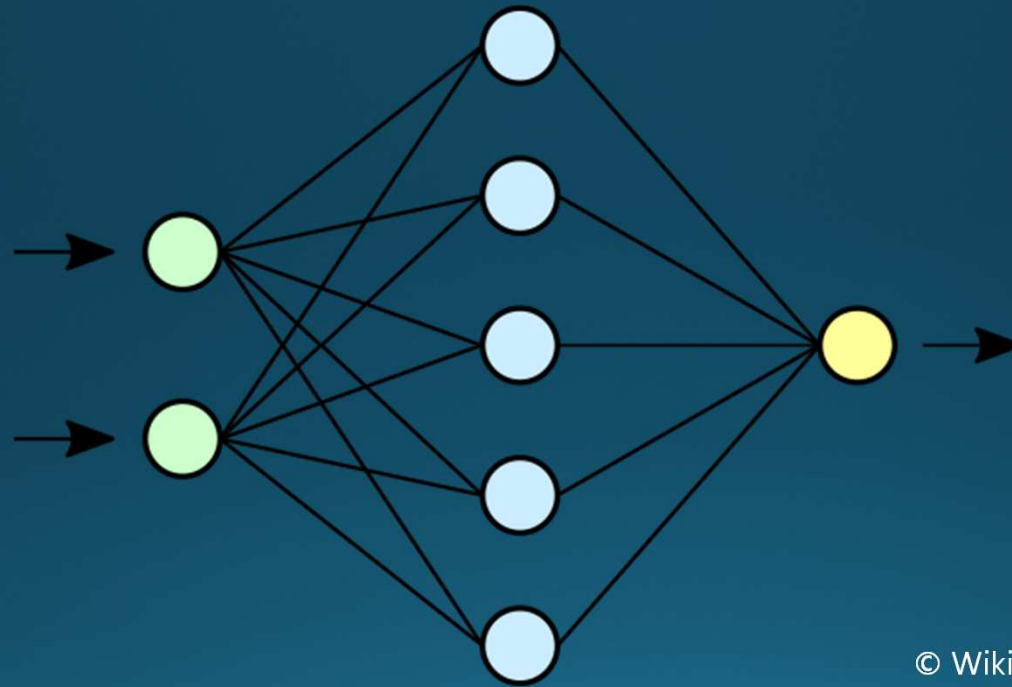
© Wikimedia



Maschinelles Lernen mit Künstlichen Neuronalen Netzen (KNN)

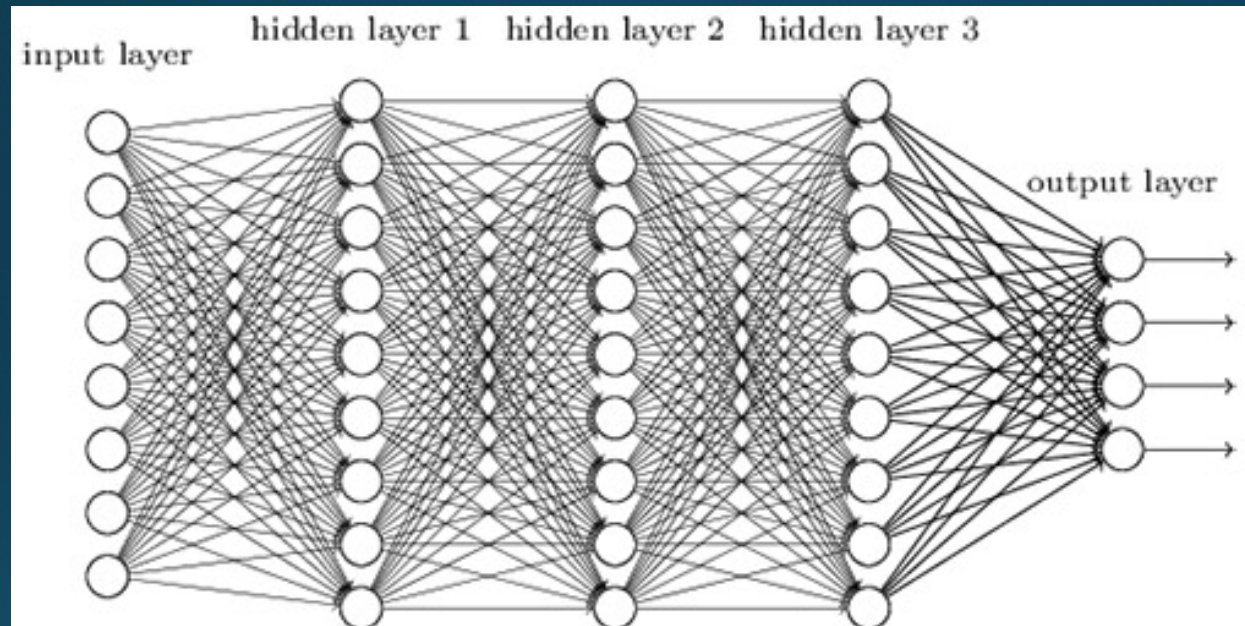
Rosenblatt, Frank (1958): *The perceptron: a probabilistic model for information storage and organization in the brain*. Psychological Reviews **65** (1958) 386-408

Künstliches Neuronales Netz



© Wikipedia

Künstliches Neuronales Netz



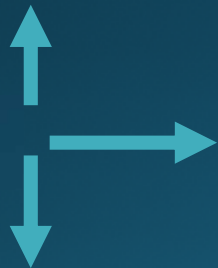
© Wikipedia

Lernverfahren

- Überwachtes Lernen



DOG



CAT



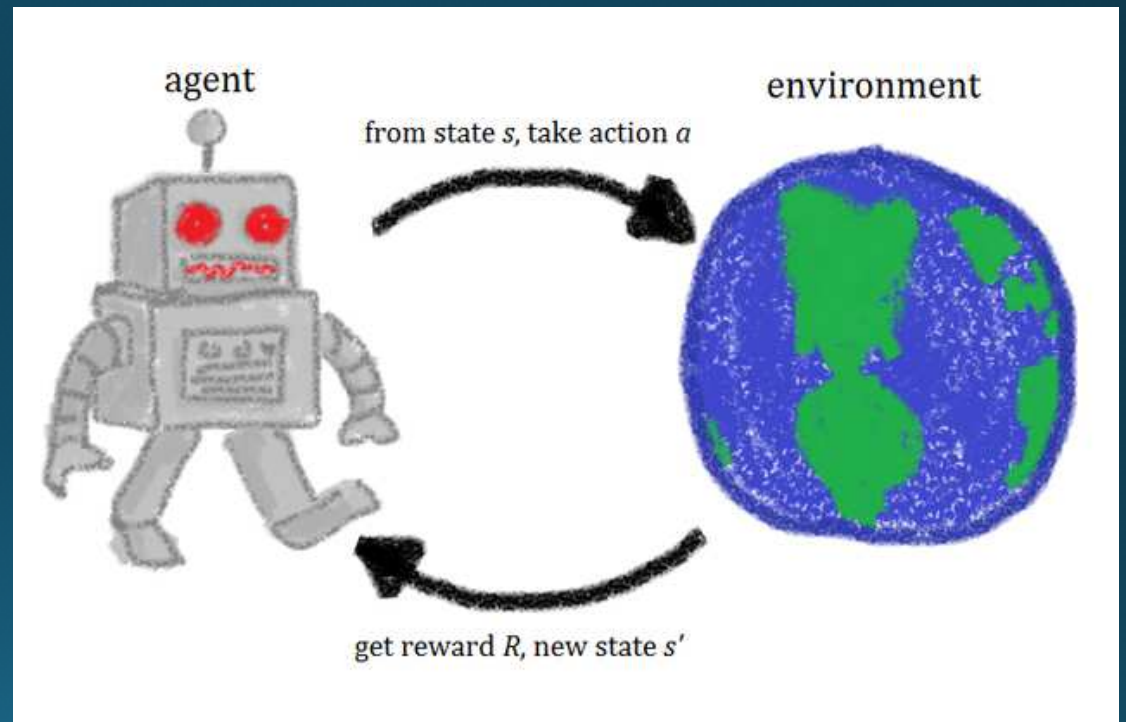
Lernverfahren

- Unüberwachtes Lernen



Lernverfahren

- Reinforcement Learning



© Wikimedia

Hybride Ansätze

- Maschinelles Lernen + Wissensbasierte Programmierung



- Hintergrundwissen durch kausale Zusammenhänge
- erklärbarer /nachvollziehbarer
- robuster
- kommen mit weniger Daten aus
- vorhandenes Wissen (etwa aus der Produktionstechnik) kann verwendet werden
- Fernes Ziel:
Gesunder Menschenverstand / Common Sense

Computer Science > Computation and Language

[Submitted on 12 Jun 2017 (v1), last revised 6 Dec 2017 (this version, v5)]

Attention Is All You Need

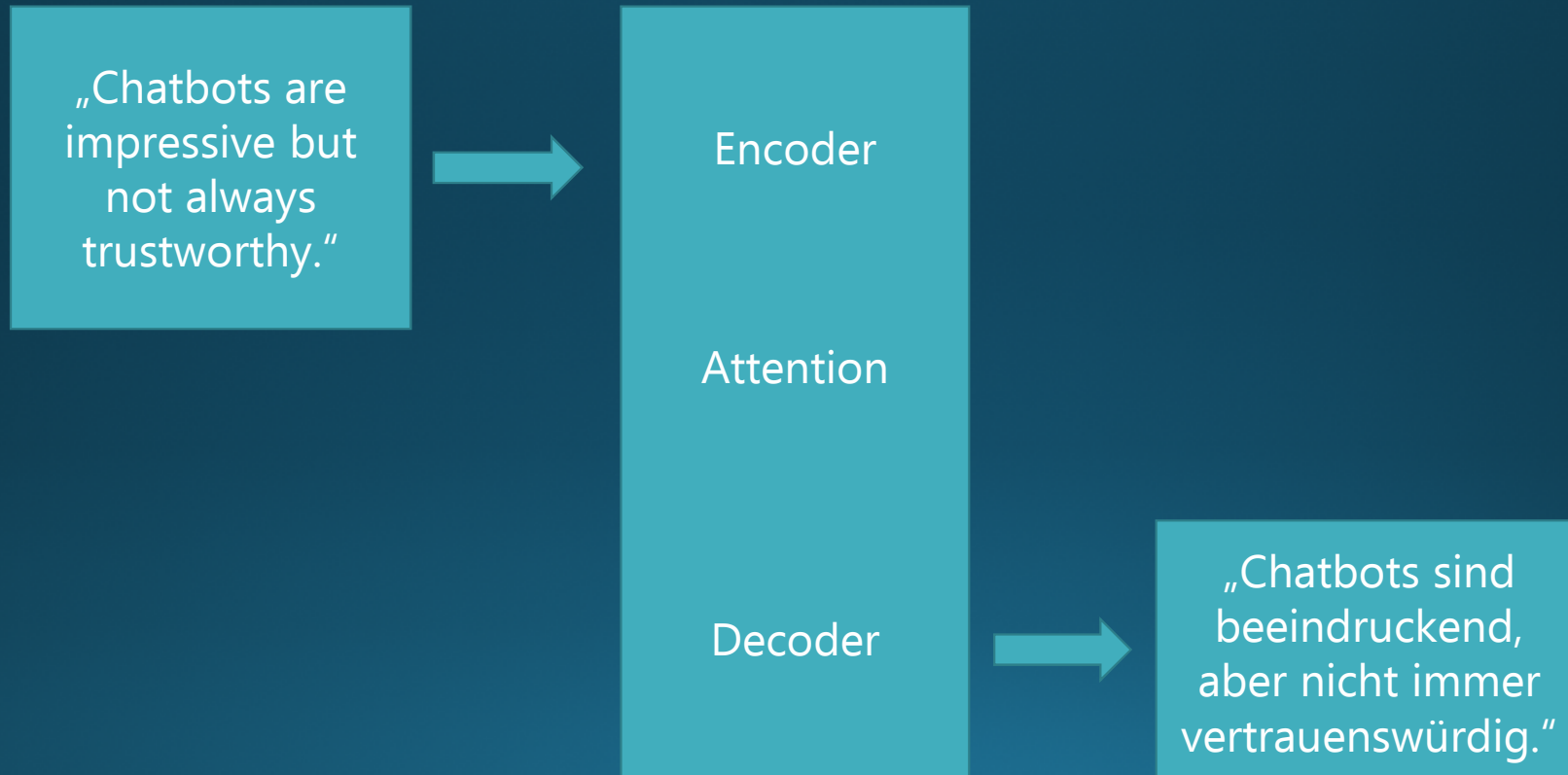
Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin

“We propose a new simple network architecture, the **Transformer**, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely.”

© Alex Goy, Creative Commons

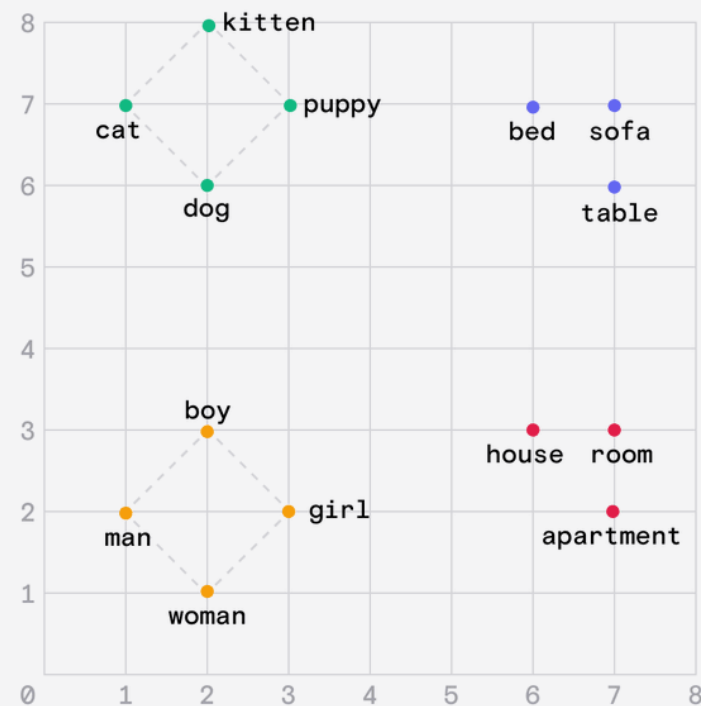


Transformer



Transformer

- Sätze zu Token
- Einbettung, „embedding“ der Token
- Aufmerksamkeit, „attention“ auf den Kontext



Mächtige Werkzeuge

Groß, größer, ...

- GPT-3 (Open AI): 175 Milliarden Parameter
- GPT-4 (Open AI): 100 Billionen?
- PanGu-Alpha (Huawei): 200 Milliarden Parameter
- Switch Transformer (Google): 1,6 Billionen Parameter
- Wu Dao 2.0 (Beijing Academy of Artificial Intelligence): 1,75 Billionen Parameter

Mächtige Werkzeuge

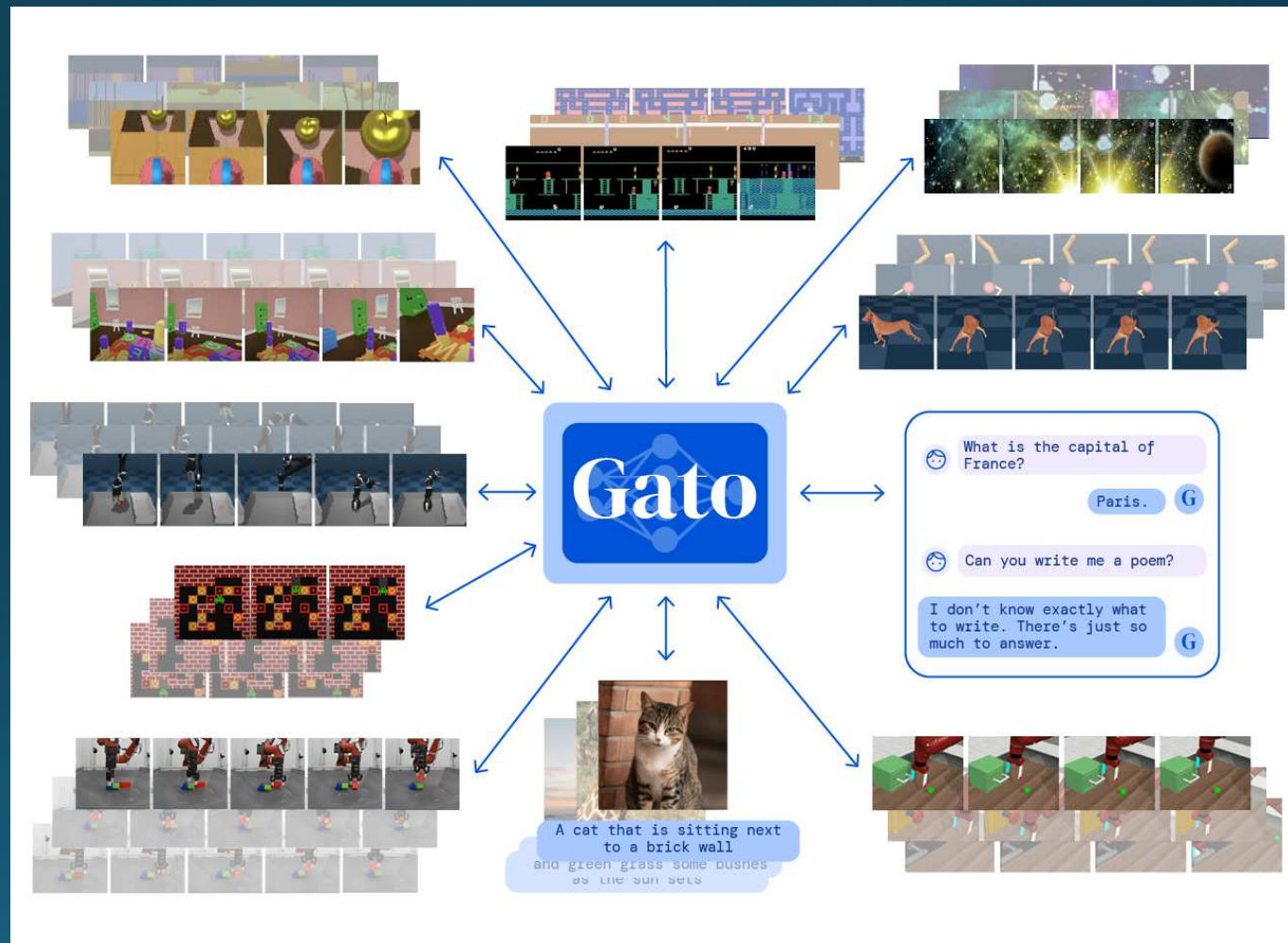
DALL-E: 12 Milliarden Parameter: Bilder zu Texten

„Ein Sessel in Form einer Avocado“



© Open AI

Gato „A Generalist Agent“





Welcome to ChatGPT

Log in with your OpenAI account to continue

Log in

Sign up

Wolken am Sommerhimmel

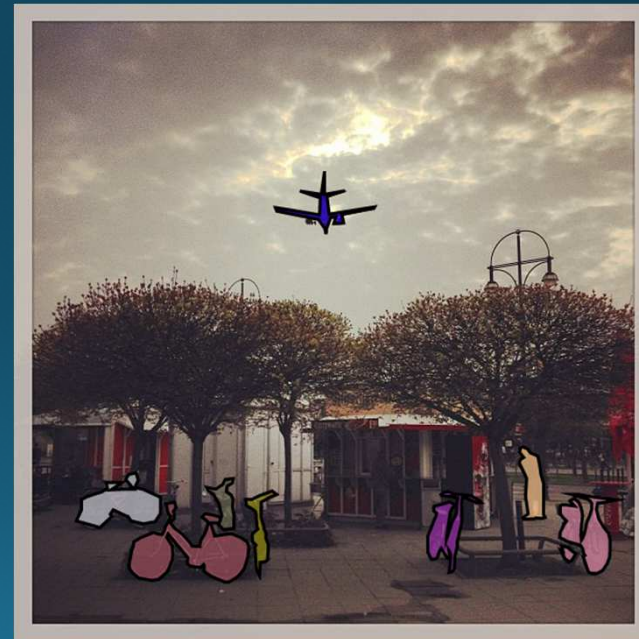
Wolken am Sommerhimmel I: Verstehen

Nell: Never Ending Language Learner
(Carnegie Mellon University 2010-2018)

- Schulgolf ist eine Sportart.
- Die „Washington Post“ ist eine Publikation.
- Salat ist ein sichtbares Objekt.
- Hektar ist eine Währung.
- Blumen sind Nahrung.
- Weniger Mais ist ein Getreide.

Wolken am Sommerhimmel II: Daten

1. Daten machen Arbeit (und das ist teuer)!
2. Nicht überall gibt es genug Daten.



COCO Dataset

<http://cocodataset.org/#home>

Wolken am Sommerhimmel III: „Vorurteile“

Machine Bias

There's software used across the country to predict future criminals. And it's biased against blacks.

*by Julia Angwin, Jeff Larson, Surya Mattu and Lauren
Kirchner, ProPublica* May 23, 2016

Wolken am Sommerhimmel IV: „Fakes“



Wolken am Sommerhimmel V: Black Box

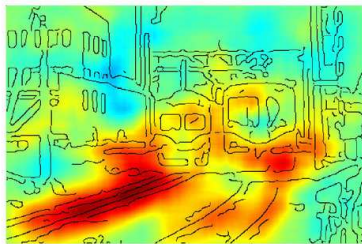


**Black-box
Algorithmus**

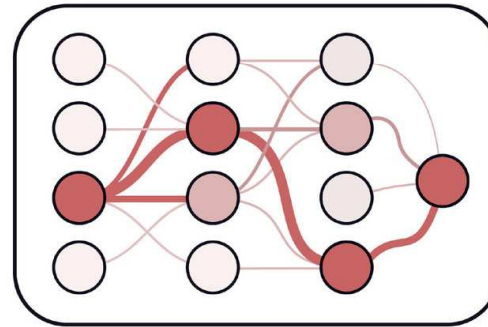
"Zug"

*Aber, worauf schaut
der Algorithmus?*

*Können wir sicher sein,
dass der Zug korrekt
erkannt wird?*



*Hier schaut der Algorithmus
NICHT auf den Zug, sondern
auf die Schienen!*



Mit der Methode des Fraunhofer HHI
und der TU Berlin kann visualisiert
werden, worauf der
black-box Algorithmus schaut

eXplainable AI (XAI)

- LIME
- Heat Maps
- Rationalisierung

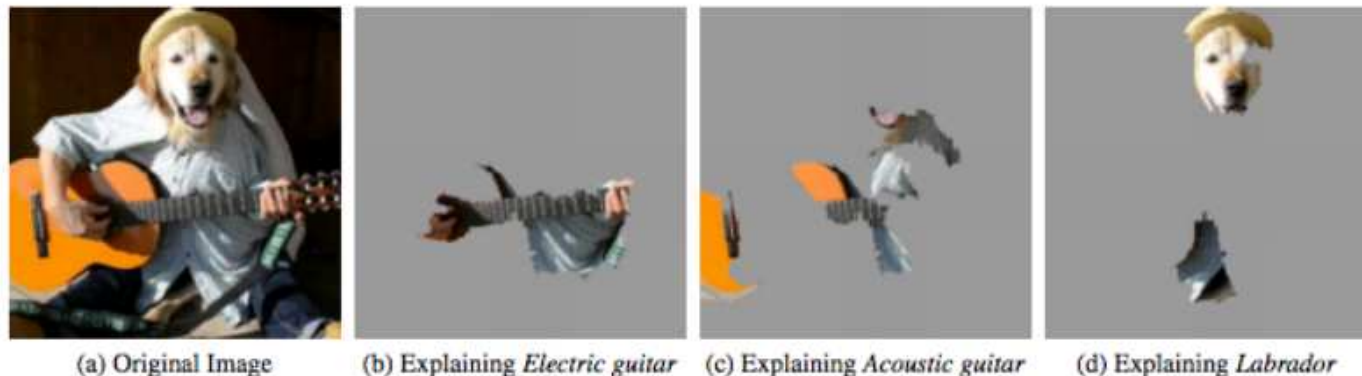


Figure 4: Explaining an image classification prediction made by Google's Inception network, highlighting positive pixels. The top 3 classes predicted are "Electric Guitar" ($p = 0.32$), "Acoustic guitar" ($p = 0.24$) and "Labrador" ($p = 0.21$)

Marco Tulio Ribeiro: LIME - Local Interpretable Model-Agnostic Explanations

Lernende Algorithmen sind problematisch in

- sicherheitskritischen Fällen
- sozialen Kontexten
- Moral und Ethik
- in militärischen Kontexten

Was tun?

Regulierung

Artificial Intelligence Act

Verordnung zur Festlegung harmonisierter Vorschriften für Künstliche Intelligenz (Gesetz über Künstliche Intelligenz)

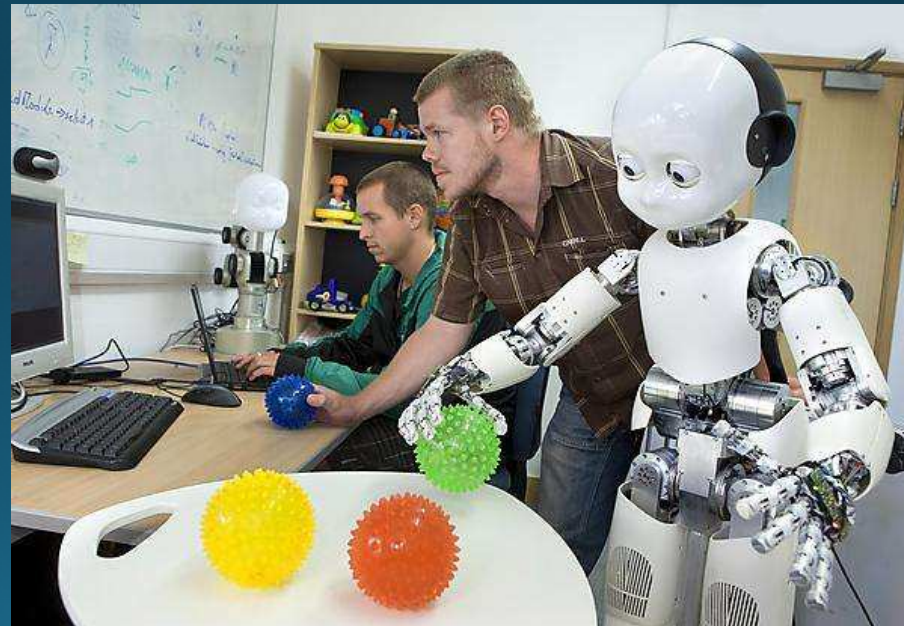
- Gilt für Anbieter und Nutzer
- Regulierung nach Risiko-Gruppen
- Verbote

Zurück zum Menschen

EU-Projekt iCub



© Univ. of Plymouth



Was wollen wir?
(Und warum?)

Vielen Dank!

www.manuela-lenzen.de

