Künstliche Intelligenz ... für Fußgänger

Jan Krieger

FORE, Frauenchiemsee, 21.7.2017

Was ist künstliche Intelligenz (KI)?

> *Artificial intelligence is the science of making machines do things that would require intelligence if done by men.*
>
> — MARVIN MINSKY, 1956

> *Artificial Intelligence is the study of mental faculties through the use of computaional models.*
>
> — EUGENE CHARNIAK & DREW MCDERMOTT, 1984

> *Artificial Intelligence is the study of the computations that make it possible to perceive, reason, and act.*
>
> — PATRICK HENRY WINSTON, 1992

> *Artificial Intelligence is a field of science and engineering concerned with the computational understanding of what is commonly called intelligent behaviour, and with the creation of artifacts (software, robots, ...) that exhibit such behaviour.*
>
> — STUART C. SHAPIRO, 1987

> *You bet. I'm a machine and you're a machine, and we both think, don't we?*
>
> — CLAUDE SHANNON

Was ist künstliche Intelligenz (KI)?

> *Artificial intelligence is the science of making machines do things that would require intelligence if done by men.*
> — Marvin Minsky, 1956

> *Artificial Intelligence is the study of mental faculties through the use of computaional models.*
> — Eugene Charniak & Drew McDermott, 1984

> *Artificial Intelligence is the study of the computations that make it possible to perceive, reason, and act.*
> — Patrick Henry Winston, 1992

> *Artificial Intelligence is a field of science and engineering concerned with the computational understanding of what is commonly called intelligent behaviour, and with the creation of artifacts (software, robots, ...) that exhibit such behaviour.*
> — Stuart C. Shapiro, 1987

> *You bet. I'm a machine and you're a machine, and we both think, don't we?*
> — Claude Shannon

- **Minsky:** Teilnehmer Dartmouth Conference (Begründung des Fachgebiets, 1956)
- **Shapiro:** sehr breite Definition
- **Charniak/McDermott:** Konzentration auf Verstehen der Gedankenprozesse
- **Winston:** Konzentration auf Produktion von KI-Artefakten

Starke KI und die Situation in den 1950ern

Definition

> *According to **weak AI**, the principal value of the computer in the study of the mind is that it gives us a very powerful tool. For example, it enables us to formulate and test hypotheses in a more rigorous and precise fashion. [...] According to **strong AI**, [...] **the appropriately programmed computer really is a mind**, in the sense that computers given the right programs can be literally said to understand [...] In strong AI, because the programmed computer has cognitive states, [...] the programs are themselves the explanation [of intelligence].*
>
> — JOHN R. SEARLE, 1980

> *The Physical Symbol System Hypothesis: A physical symbol system has the necessary and sufficient means of general intelligent action.*
>
> — NEWELL & SIMON, 1976

- Ziel: eine (echte) Intelligenz zu erschaffen, die das menschliche Denken mechanisieren soll
- These: Intelligenz ist reine Informationsverarbeitung.
- Das Ziel wirft drei große Fragen auf:
  1. Warum sollte das möglich sein?
  2. Was ist Intelligenz?
  3. Wie/mit welchen Methoden kann man KI erreichen?
- Gegenstück **schwache KI:** Mittel zum Verständnis von Intelligenz, Nachbilden von Intelligenz, keine „echte“ Intelligenz

Ref: *Newell and Simon [1976], Searle [1980]*

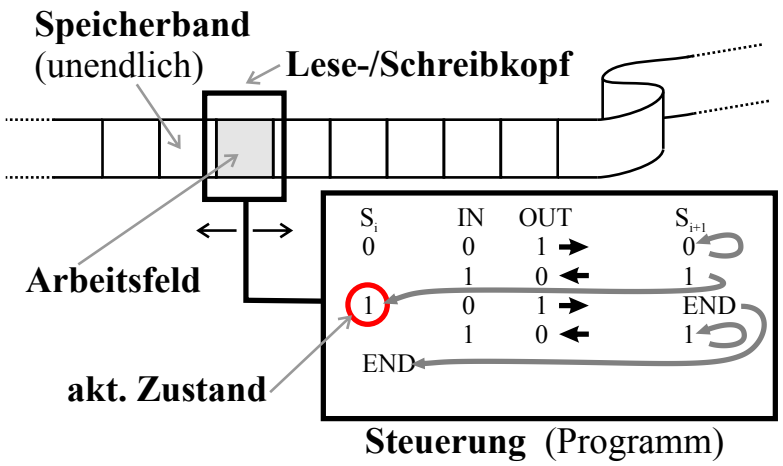
Starke KI und die Situation in den 1950ern

Stand der Informatik/KI in den 1950ern

- David Hilbert/Kurt Gödel/Bertrand Russel/...: Formalisierung der Mathematik, Möglichkeit der automatischen Beweisbarkeit von Sätzen (ab 1900)
- Einsatz von „Turing-Welchman-Bombe“ zum Knacken der deutschen Enigma (ab 1940)
- Alan Turing: Allgemeine Theorie der Berechenbarkeit und von „Computern“ (1936)
- These von Alan Turing & Alonzo Church: *Jede „intuitiv“ berechenbare Funktion ist auch von einer Turing-Maschine berechenbar.*

MISSING IMAGE

Quelle: https://de.wikipedia.org/wiki/Turing-Bombe

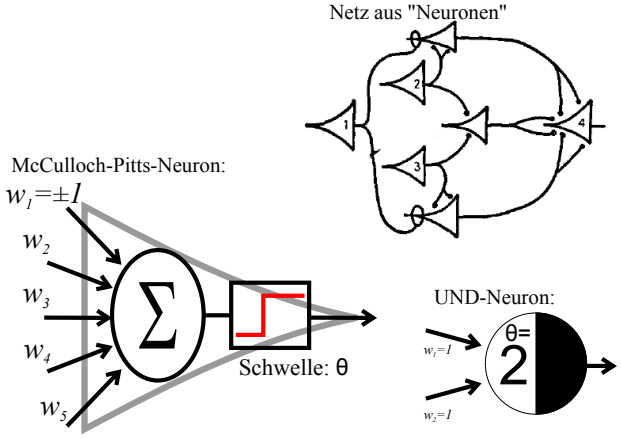


Ref: *Whitehead and Russel [1963], Turing [1950]*

Starke KI und die Situation in den 1950ern

Stand der Informatik/KI in den 1950ern

- Warren S. McCulloch & Walter Pitts mathemtisches Modell für Neuronen (1943) und zeigen, dass damit gerechnet werden kann (aber: nicht so mächtig, wie Turing-Maschine)
- Frank Rosenblatt: erweitertes Modell „Perceptron“ (1956): zur Mustererkennung/Klassifikation verwendet werden (aber nur linear separierbare Muster)
- Noam Chomsky veröffentlicht „Syntactic Structures“ über formale Sprachen (1956),
  $\rightarrow$ Wissensrepräsentation

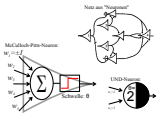


Ref: *McCulloch and Pitts [1943], Rosenblatt [1958]*

Starke KI und die Situation in den 1950ern

Stand der Informatik/KI in den 1950ern

- Warren S. McCulloch & Walter Pitts mathemtisches Modell für Neuronen (1943) und zeigen, dass damit gerechnet werden kann (aber: nicht so mächtig, wie Turing-Maschine)
- Frank Rosenblatt: erweitertes Modell „Perceptron“ (1956): zur Mustererkennung/Klassifikation verwendet werden (aber nur linear separierbare Muster)
- Noam Chomsky veröffentlicht „Syntactic Structures“ über formale Sprachen (1956), → Wissensrepräsentation



Ref: *McCulloch and Pitts [1943], Rosenblatt [1958]*

- Diese Befunde haben dazu geführt, anzunehmen, dass Computer zu intelligentem Verhalten fähig sein könnten
- Warren S. McCulloch & Walter Pitts:
  - Eingang: 1 oder 0, fixe Schwelle, Ausgang 1, 0, Gewichte -1,0,+1
  - Kann alle logischen Funktionen nachbilden, da sich AND/OR/NOT aus McCulloch-Pitts-Zellen bauen lassen. Diese genügen eigentlich um alle logischen (Bool'schen) Funktionen aufzubauen.
  - ABER: M. Minsky zeigte, dass sich XOR (exklusives ODER: 1-0, 0-1 führen zu Ausgabe 1, aber nicht 1-1) trotzdem nicht bauen lässt!!!
  - Vernetzung erlaubt kompliziertere Funktionen zu implementieren
- Perzeptron: Erweitertes Neuronen-Modell
- Chomsky's Arbeiten über formale Sprachen erlauben es Informationen für einen Computer verständlich zu kodieren. Sie bilden so eine der Basisarbeiten (zusammen mit TM und Shannon's Arbeiten zur Informationsübertragung und -codierung) der (theoretischen) Informatik.
- Alle diese Befunde ließen Anfang der 1950er Jahre den Versuch eine KI zu konstruieren möglich und sinnvoll erscheinen:
  - Church-Turing-These: Computer können viel (wenn nicht alles) von dem berechnen, was Menschen berechnen können
  - McCulloch-Pitts-Neuronen: Struktur des Gehirns kann (krude) modelliert werden und dies erlaubt schon viele Berechnungen $\longrightarrow$ d.h. kann man mit

Starke KI: Was ist Intelligenz?

- Intelligenz ist, was ein **IQ-Test** misst
- **Spearman's Generalfaktor der Intelligenz**:
  führt die positiv korrelierenden Ergebnisse verschiedener Intelligenztests auf einen Faktor zurück.
- Howard Gardeners **multiple Intelligenzen**:
  - *Sprachlich-Linguistische Intelligenz*
    Sprache gebrauchen, nutzen, lernen ...
  - *Logisch-mathematische Intelligenz*
    Probleme logisch analysieren, mathematische Fähigkeiten, wissenschaftliche Methoden anwenden ...
  - *Musikalisch-rhythmische Intelligenz*
    Musizieren, Komponieren, Sinn für Musik ...
  - *Bildlich-räumliche Intelligenz*
    Erfassen weiter und begrenzter Räume ...
  - *Körperlich-kinästhetische Intelligenz*
    Einsatz/Steuerung des eigenen Körpers, um Ziele zu erreichen
  - *Naturalistische Intelligenz*
    Sensibilität für Naturphänomene, Beobachtung, ...
  - *Interpersonale/Soziale Intelligenz*
    Empathie, Beinflussung der Stimmung/Emotionen anderer ...
  - *Intrapersonelle Intelligenz*
    eigene Gefühle, Stimmungen, Schwächen, Antriebe und Motive verstehen und beeinflussen ...

Ref: *Spearman [1904], Gardner [2011]*

Starke KI: Was ist Intelligenz?

- Louis Leon Thurstone's **sieben Primärfaktoren**:
  - *S (space):*
    räumlich-visuelle Aufgaben
  - *P (perceptual speed):*
    Wahrnehmung von Objekten und Relation zwischen ihnen
  - *N (numerical ability):*
    rechnerisch-mathematische Fähigkeiten
  - *M (memory):*
    Gedächtnisleistung
  - *R (reasoning):*
    logisches Schlussfolgern
  - *W (word fluency):*
    Wortflüssigkeit
  - *V (verbal relations):*
    verbale Beziehung richtig verstehen und interpretieren

Starke KI: Was ist Intelligenz?

- Louis Leon Thurstone's **sieben Primärfaktoren**:
  - *S (space):* räumlich-visuelle Aufgaben
  - *P (perceptual speed):* Wahrnehmung von Objekten und Relation zwischen ihnen
  - *N (numerical ability):* rechnerisch-mathematische Fähigkeiten
  - *M (memory):* Gedächtnisleistung
  - *R (reasoning):* logisches Schlussfolgern
  - *W (word fluency):* Wortflüssigkeit
  - *V (verbal relations):* verbale Beziehung richtig verstehen und interpretieren

- Spearman's Generalfaktor der Intelligenz:
  - ... führt die positiv korrelierenden Ergebnisse verschiedener Intelligenztests auf einen Faktor zurück (statistisches Maß aus der Faktoranalyse).
  - ... hat hohe Vorhersagekraft für späteren Erfolg (Beruf, Schule, ...) im Leben, bei Messung in Kindheit
- **D.h. Es gibt viele verschiedene Intelligenzmodelle, aber keine definite Definition, was menschliche Intelligenz ist oder ausmacht. Aber es gibt Ideen, welche Fähigkeiten ein intelligenter Mensch hat und diese können als Startpunkt für KI-Entwicklung genutzt werden.**

Starke KI: Was ist Intelligenz?

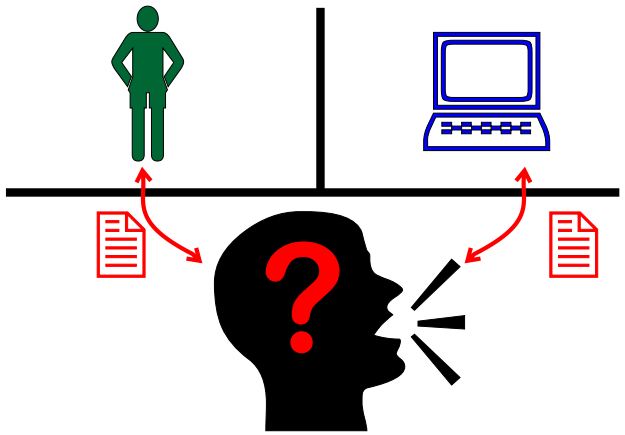
Anforderungen an künstliche Intelligenz

- Wissenrepräsentation, umgehen mit unscharfen und fehlenden Informationen, Unsicherheiten
- Problemlösestrategien finden, Schlussfolgern (Suche)
- Planen (Vorhersagen)
- maschinelles Lernen
- evtl. natürliche Sprache verstehen und sprechen
- evtl. Wahrnehmung der Welt über Sensoren
- evtl. Interaktion mit der Welt über Aktoren

Starke KI: Was ist Intelligenz?

TURING-Test

- Ersetze die Frage „Können Maschinen Denken?“ durch ein Spiel, den TURING-Test:
  - Eine Person soll durch schriftliche Fragen/Antworten herausfinden, welcher von zwei Probanden ein Mensch und welcher eine Maschine ist.
  - Computer wäre also von einem Menschen nicht zu unterscheiden.
- benötigte Fähigkeiten:
  - Sprache, Fragen verstehen (auch „zwischen den Zeilen“)
  - Antworten formulieren
  - Wissen, umgehen mit Nicht-Wissen
  - Schlussfolgern
  - Empathie für Frager, Anpassen/Lernen



Ref: *Turing [1950]*

Starke KI: Was ist Intelligenz?

TURING-Test

- Ersetze die Frage „Können Maschinen Denken?“ durch ein Spiel, den TURING-Test:
  - Eine Person soll durch schriftliche Fragen/Antworten herausfinden, welcher von zwei Probanden ein Mensch und welcher eine Maschine ist.
  - Computer wäre also von einem Menschen nicht zu unterscheiden.
- benötigte Fähigkeiten:
  - Sprache, Fragen verstehen (auch „zwischen den Zeilen“)
  - Antworten formulieren
  - Wissen, umgehen mit Nicht-Wissen
  - Schlussfolgern
  - Empathie für Frager, Anpassen/Lernen

Ref: *Turing [1950]*

- bisher kein KI-Programm, das den TURING-Test besteht
- fraglich, ob TURING-Test wirklich Intelligenz misst
- evtl. sucht man nur eine gute Täuschung ... aber ist eine Täuschung nicht Realität, wenn sie von dieser nicht zu unterscheiden ist?

Starke KI: Joseph Weizenbaum's ELIZA (1966)

- erster Chatter-Bot
- arbeitet mit einfachen Satz-Zerlegungen, Thesaurus und Ersetzungsregeln

```
U: Men are all alike.
E: IN WHAT WAY?
U: They're always bugging us about something or other.
E: CAN YOU THINK OF A SPECIFIC EXAMPLE?
U: my boyfriend made me come here.
E: YOUR BOYFRIEND MADE YOU COME HERE?
U: He says I'm  depressed much of the time.
E: I AM SORRY TO HEAR YOU ARE DEPRESSED.
...
U: Perhaps I could learn to get along with my mother.
E: TELL ME MORE ABOUT YOUE FAMILY.
...
U: My father is afraid of everybody.
E: WHAT ELSE COMES TO MIND WHEN YOU THINK OF YOUR FATHER?
U: Bullies.
E: DOES THAT HAVE ANYTHING TO DO WITH THE FACT THAT YOUR
BOYFRIEND MADE YOU COME HERE?
```

Ref: *Weizenbaum [1966]*

Starke KI: Joseph Weizenbaum's ELIZA (1966)

- erster Chatter-Bot
- arbeitet mit einfachen Satz-Zerlegungen, Thesaurus und Ersetzungsregeln

U: *Men are all* *alike* (1) .
E: IN WHAT WAY?
U: *They're* *always* (1) *bugging us about something or other.*
E: CAN YOU THINK OF A SPECIFIC EXAMPLE?
U: *my boyfriend made me come here.*
E: YOUR BOYFRIEND MADE YOU COME HERE?
U: *He says I'm depressed much of the time.*
E: I AM SORRY TO HEAR YOU ARE DEPRESSED.
...
U: *Perhaps I could learn to get along with my mother.*
E: TELL ME MORE ABOUT YOUE FAMILY.
...
U: *My father is afraid of everybody.*
E: WHAT ELSE COMES TO MIND WHEN YOU THINK OF YOUR FATHER?
U: *Bullies.*
E: DOES THAT HAVE ANYTHING TO DO WITH THE FACT THAT YOUR BOYFRIEND MADE YOU COME HERE?

Regeln:

„alike"
→ „In what way?"
→ „What resemblence do you see ?"
→ ...

„always"
→ „Can you think of a specific example ?"
→ „When?"
→ ...

Ref: *Weizenbaum [1966]*

Starke KI: Joseph Weizenbaum's ELIZA (1966)

- erster Chatter-Bot
- arbeitet mit einfachen Satz-Zerlegungen, Thesaurus und Ersetzungsregeln

U: *Men are all alike.*
E: IN WHAT WAY?
U: *They're always bugging us about something or other.*
E: CAN YOU THINK OF A SPECIFIC EXAMPLE?
U: *Well, my (1) boyfriend made me come here (2) .*
E: YOUR BOYFRIEND MADE YOU COME HERE?
U: *He says I'm depressed much of the time.*
E: I AM SORRY TO HEAR YOU ARE DEPRESSED.
...
U: *Perhaps I could learn to get along with my mother.*
E: TELL ME MORE ABOUT YOUE FAMILY.
...
U: *My father is afraid of everybody.*
E: WHAT ELSE COMES TO MIND WHEN YOU THINK OF YOUR FATHER?
U: *Bullies.*
E: DOES THAT HAVE ANYTHING TO DO WITH THE FACT THAT YOUR BOYFRIEND MADE YOU COME HERE?

Regeln:

„my"
↪ „* my *"
→ „Your (2) ?"
→ „Is it important to you that your (2)?"
→ ...

Ref: *Weizenbaum [1966]*

Starke KI: Joseph Weizenbaum's ELIZA (1966)

- erster Chatter-Bot
- arbeitet mit einfachen Satz-Zerlegungen, Thesaurus und Ersetzungsregeln

U: *Men are all alike.*
E: IN WHAT WAY?
U: *They're always bugging us about something or other.*
E: CAN YOU THINK OF A SPECIFIC EXAMPLE?
U: *my boyfriend made me come here.*
E: YOUR BOYFRIEND MADE YOU COME HERE?
U: *He says* *I'm* (1) (2) *depressed* (3) *much of the time.*
E: I AM SORRY TO HEAR YOU ARE DEPRESSED.
...
U: *Perhaps I could learn to get along with my mother.*
E: TELL ME MORE ABOUT YOUE FAMILY.
...
U: *My father is afraid of everybody.*
E: WHAT ELSE COMES TO MIND WHEN YOU THINK OF YOUR FATHER?
U: *Bullies.*
E: DOES THAT HAVE ANYTHING TO DO WITH THE FACT THAT YOUR BOYFRIEND MADE YOU COME HERE?

Thesaurus:

„depressed"/"unhappy"/... → „sad"
„I'm"/... → „I am"

Regeln:

„I"
↪ „* i am* @sad *"
→ „I am sorry to hear that you are (3)."
→ „Can you explain what made you (3) ?"
→ ...

Ref: *Weizenbaum [1966]*

Starke KI: Joseph Weizenbaum's ELIZA (1966)

- erster Chatter-Bot
- arbeitet mit einfachen Satz-Zerlegungen, Thesaurus und Ersetzungsregeln

U: *Men are all alike.*
E: IN WHAT WAY?
U: *They're always bugging us about something or other.*
E: CAN YOU THINK OF A SPECIFIC EXAMPLE?
U: *my boyfriend made me come here* (MEM) .
E: YOUR BOYFRIEND MADE YOU COME HERE?
U: *He says I'm depressed much of the time.*
E: I AM SORRY TO HEAR YOU ARE DEPRESSED.
...
U: *Perhaps I could learn to get along with my mother.*
E: TELL ME MORE ABOUT YOUE FAMILY.
...
U: *My father is afraid of everybody.*
E: WHAT ELSE COMES TO MIND WHEN YOU THINK OF YOUR FATHER?
U: *Bullies* (1) .
E: DOES THAT HAVE ANYTHING TO DO WITH THE FACT THAT YOUR BOYFRIEND MADE YOU COME HERE?

(MEM): Gedächtnis:
„my"
→ **Regel anwenden & speichern:**
↪ „$ * my *"
→ „Does that have anything to do with the fact that your (2) ?"
→ „Lets discuss further why your (2)."
→ ...

Regeln:
→ KEIN TREFFER!
↪ hole aus Gedächtnis

Ref: *Weizenbaum [1966]*

Starke KI: Joseph Weizenbaum's ELIZA (1966)

- erster Chatter-Bot
- arbeitet mit einfachen Satz-Zerlegungen, Thesaurus und Ersetzungsregeln

```
U: Men are all alike.
E: IN WHAT WAY?
U: They're always bugging us about something or other.
E: CAN YOU THINK OF A SPECIFIC EXAMPLE?
U: my boyfriend made me come here .
E: YOUR BOYFRIEND MADE YOU COME HERE?
U: He says I'm  depressed much of the time.
E: I AM SORRY TO HEAR YOU ARE DEPRESSED.
...
U: Perhaps I could learn to get along with my mother.
E: TELL ME MORE ABOUT YOUR FAMILY.
...
U: My father is afraid of everybody.
E: WHAT ELSE COMES TO MIND WHEN YOU THINK OF YOUR FATHER?
U: Bullies .
E: DOES THAT HAVE ANYTHING TO DO WITH THE FACT THAT YOUR
BOYFRIEND MADE YOU COME HERE?
```

(MEM): Gedächtnis:

```
„my"
  ↳ Regel anwenden & speichern:
      ↳ „* my *"
        → „Does that have anything to do with the
            fact that your (2) ?"
        → „Lets discuss further why your (2)."
        → ...

Regeln:
  → KEIN TREFFER!
      ↳ hole aus Gedächtnis
```

Ref: *Weizenbaum [1966]*

- Chater-Bot = Programm, dass eine Konverstaion mit einem (menschlichen) Nutzer führt, wird heute z.B. bei Nutzer-Anfragen in Online-Shops benutzt
- ein einfaches Programm ist schon nicht allzu-schlecht im führen einer Konversation
- Weizenbaum installierte sein Programm auf dem Mainframe seiner Universität und war überrascht, wie viele Menschen es für Unterhaltungen nutzten
- Kurzzeitig wurde auch diskutiert solche Programme in der Psychotherapie zu nutzen

Starke KI: Erste Erfolge

MISSING IMAGE

http://www.computerhistory.org/collections/catalog/102630790

- Claude Shannon baut erste Schach-spielende Maschine (1949)
- Claude Shannon eine Maschine mit Gedächtnis, die Labyrinthe löst (1952)
- Erster Sieg eine Schachprogramms über Menschen (1956),
  aber mit vereinfachten Regeln und 6×6-Brett
- Allen Newell, John Simon & Herbert Shaw: „General Problem Solver“ (1956)
  kann einfache mathematische Beweise führen, indem er von einem Satz von Axiomen und erlaubten Regeln (Schlussfolgerungsmöglichkeiten) ausgeht und eine gültige Abfolge dieser anwendet (=Schlussfolgerungskette des Beweises), um zur gewünschten Behauptung zu gelangen.

$\Rightarrow$ **Generell: Zunächst Konzentration auf symbolische Systeme, regel-basierte Systeme**

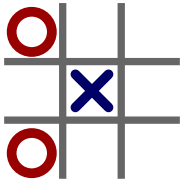
Ref: *Shannon [1950], Newell et al. [1959], von Neumann [1928, 1945]*

Frühe starke KI: Kombinatorische Explosion

- bei Spiele-KI/Beweis-Sucher: Bei jedem Zug muss eine optimale Wahl zwischen verschiedenen Alternativen getroffen werden.
- z.B. Tic-Tac-Toe: max. 9 Züge, Anzahl Möglichkeiten: $9 \cdot 8 \cdot 7 \cdot \ldots \cdot 1 = 362880$
- Schach/Dame: viel mehr Möglichkeiten $\rightarrow$ nutze Heuristiken ... oder schnellere Computer!

Frühe starke KI: Kombinatorische Explosion

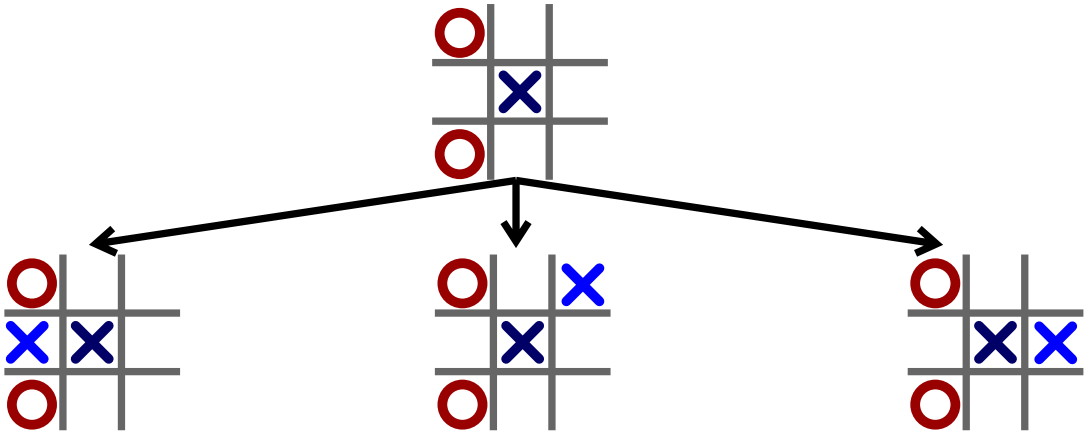
- bei Spiele-KI/Beweis-Sucher: Bei jedem Zug muss eine optimale Wahl zwischen verschiedenen Alternativen getroffen werden.
- z.B. Tic-Tac-Toe: max. 9 Züge, Anzahl Möglichkeiten: $9 \cdot 8 \cdot 7 \cdot \ldots \cdot 1 = 362880$



- Schach/Dame: viel mehr Möglichkeiten $\rightarrow$ nutze Heuristiken ... oder schnellere Computer!

Frühe starke KI: Kombinatorische Explosion

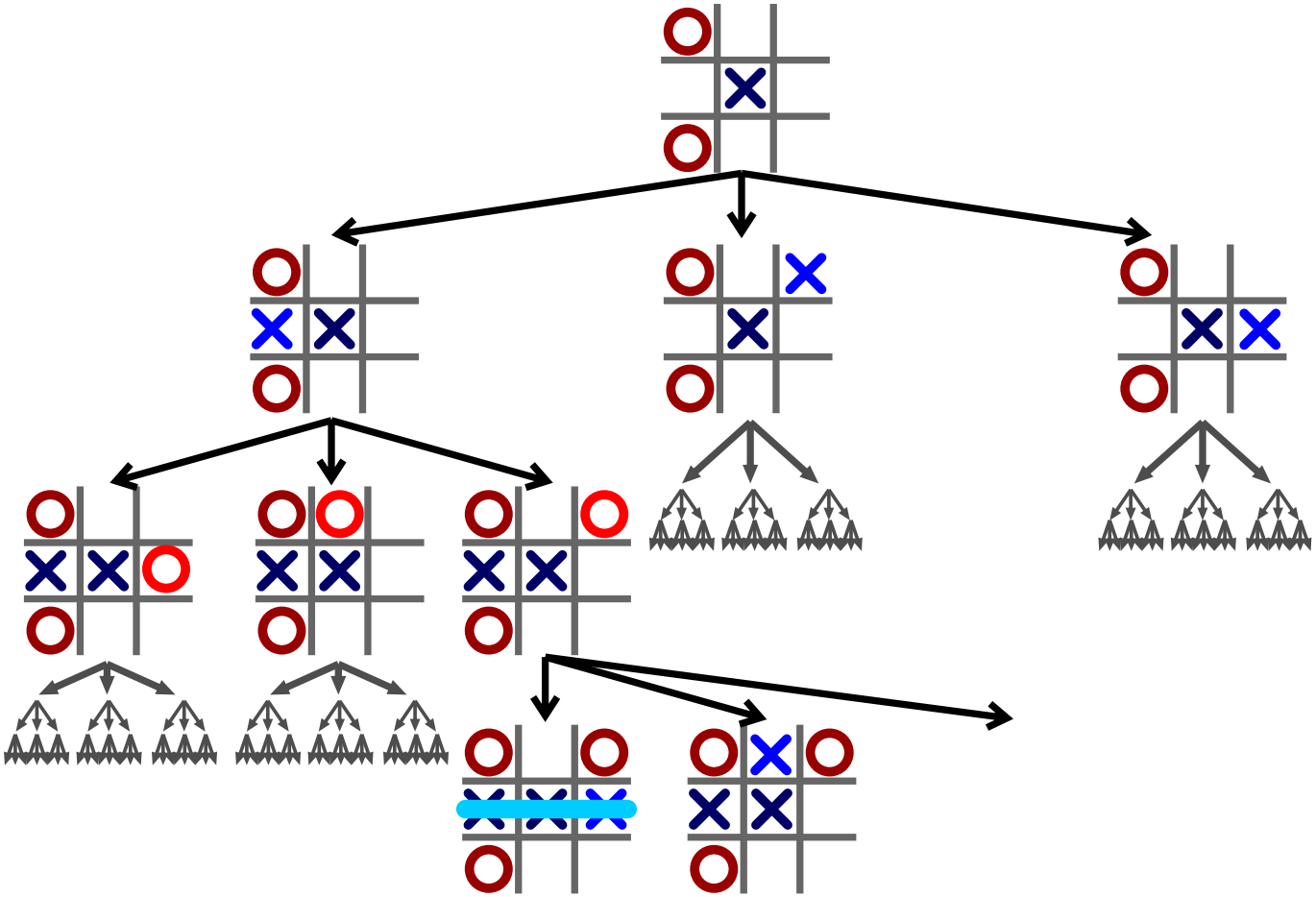
- bei Spiele-KI/Beweis-Sucher: Bei jedem Zug muss eine optimale Wahl zwischen verschiedenen Alternativen getroffen werden.
- z.B. Tic-Tac-Toe: max. 9 Züge, Anzahl Möglichkeiten: $9 \cdot 8 \cdot 7 \cdot \ldots \cdot 1 = 362880$



- Schach/Dame: viel mehr Möglichkeiten $\rightarrow$ nutze Heuristiken ... oder schnellere Computer!

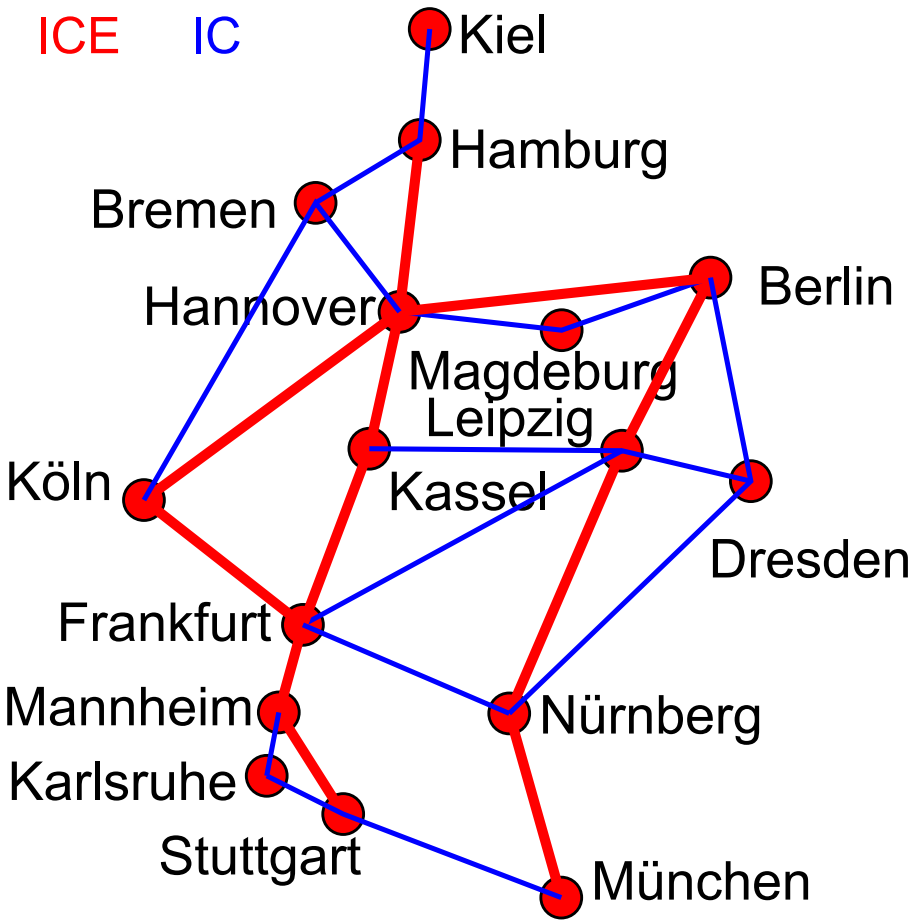
Frühe starke KI: Kombinatorische Explosion

- bei Spiele-KI/Beweis-Sucher: Bei jedem Zug muss eine optimale Wahl zwischen verschiedenen Alternativen getroffen werden.
- z.B. Tic-Tac-Toe: max. 9 Züge, Anzahl Möglichkeiten: $9 \cdot 8 \cdot 7 \cdot \ldots \cdot 1 = 362880$



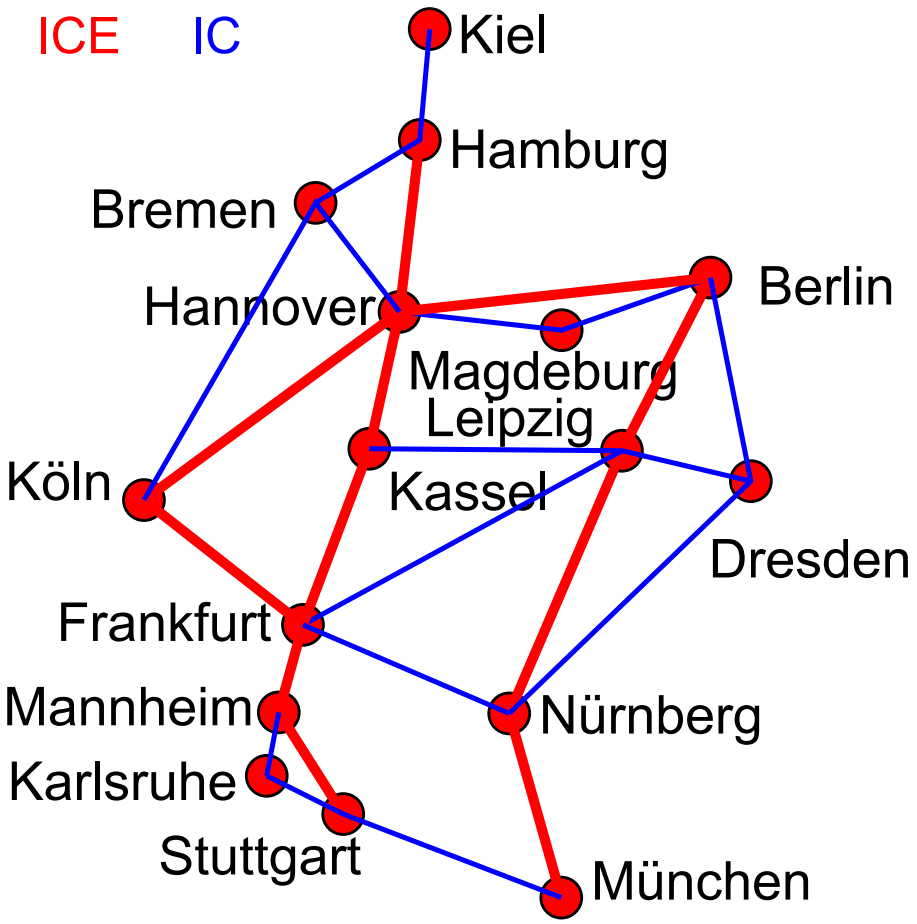
- Schach/Dame: viel mehr Möglichkeiten $\longrightarrow$ nutze Heuristiken ... oder schnellere Computer!

Frühe starke KI: Kombinatorische Explosion



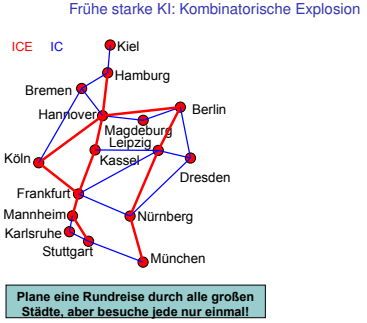
Wie kommt man am schnellsten von Hamburg nach München?

Frühe starke KI: Kombinatorische Explosion



Plane eine Rundreise durch alle großen Städte, aber besuche jede nur einmal!

Frühe starke KI: Kombinatorische Explosion



- Die ANzahl der Möglichkeiten in vielen Probleme (z.B. Spiele wie Schach/Dame/Tic-TacToe) steigt exponentiell mit der Größe des Problems (d.h. z.B. des Spielfelds, der Anzahl der Züge ...), d.h. eine Verdopplung der Spielfeldgröße führt zu einer Vervierfachung ($= 4 = 2^4$) der Möglichkeiten, eine Verdreifachung der Spielfeldgröße zu einer Verachtfachung ($= 8 = 2^3$) der Möglichkeiten ...
- Ein anderes Beispiel für dieses Problem ist es die optimale (schnellste, bequemste, günstigste, ...) Route zwischen z.B. München und Hamburg zu finden
- Eine Erweiterung ist das Handelsreisendenproblem („traveling salesman problem“), bei dem eine Anzahl von Städten besucht werden soll, wobei jede Stadt nur einmal betreten werden soll und man in der Ursprungsstadt wieder am Ende anlagt. Das Problem die kürzeste solche Route zu finden ist exponentiell (in der Anzahl der Städte) und gehört zu den sog. NP-harten Problemen (wie auch das Knacken aller heute verwendeter Verschlüsselungsverfahren)
- Um solche Probleme bei größeren Problemgrößen noch lösen zu können, verwendet man heute oft Heuristiken, die bestimmte Teile des Möglichkeitenbaums aufgrund von informiertem Raten ausschließen. Sie garantieren dann keine optimale Lösung, aber oft eine (für praktische Anwendungen) gut-genuge Lösung.

Starke KI in der Krise (1974-1980)

- erste KI-Winter (1974-1980): KI-Forschung bleibt nach frühen Erfolgen stecken auf dem Weg zur starken KI, keine Finanzierung mehr
- einige Gründe:
  - **kombinatorische Explosion** und **niedrige Rechenleistung**
  - Unmöglichkeit ausreichender **Wissenskodierung** (Menge, Qualität)
  - **Philosophische Unsicherheit**: Gödel zeigt, dass es in jedem mathematisch-logischem System Sätze gibt, die zwar wahr sind, aber nicht beweisbar.
    $\Rightarrow$ Ist das menschliche Gehirn also mächtiger als ein Computer?
  - **Moravec's Paradoxon**: logisches Schließen ist für Computer deutlich einfacher, als Spracherkennung oder Bildverstehen

  $\Rightarrow$ Top-Down-Ansatz hat nicht funktioniert

> *Encoded in the large, highly evolved sensory and motor portions of the human brain is a billion years of experience about the nature of the world and how to survive in it. The deliberate process we call reasoning is, I believe, the thinnest veneer of human thought, effective only because it is supported by this much older and much more powerful, though usually unconscious, sensorimotor knowledge. We are all prodigious olympians in perceptual and motor areas, so good that we make the difficult look easy. Abstract thought, though, is a new trick, perhaps less than 100 thousand years old. We have not yet mastered it. It is not all that intrinsically difficult; it just seems so when we do it.*
>
> — HANS MORAVEC, 1988

Ref: *Gödel [1931], Moravec [1990]*

Starke KI in der Krise (1974-1980)

- erste KI-Winter (1974-1980): KI-Forschung bleibt nach frühen Erfolgen stecken auf dem Weg zur starken KI, keine Finanzierung mehr
- einige Gründe:
  - **kombinatorische Explosion** und **niedrige Rechenleistung**
  - Unmöglichkeit ausreichender **Wissenskodierung** (Menge, Qualität)
  - **Philosophische Unsicherheit**: Gödel zeigt, dass es in jedem mathematisch-logischem System Sätze gibt, die zwar wahr sind, aber nicht beweisbar.
    ⇒ Ist das menschliche Gehirn also mächtiger als ein Computer?
  - **Moravec's Paradoxon**: logisches Schließen ist für Computer deutlich einfacher, als Spracherkennung oder Bildverstehen

  ⇒ Top-Down-Ansatz hat nicht funktioniert

> *Encoded in the large, highly evolved sensory and motor portions of the human brain is a billion years of experience about the nature of the world and how to survive in it. The deliberate process we call reasoning is, I believe, the thinnest veneer of human thought, effective only because it is supported by this much older and much more powerful, though usually unconscious, sensorimotor knowledge. We are all prodigious olympians in perceptual and motor areas, so good that we make the difficult look easy. Abstract thought, though, is a new trick, perhaps less than 100 thousand years old. We have not yet mastered it. It is not all that intrinsically difficult; it just seems so when we do it.*
>
> — HANS MORAVEC, 1988

Ref: *Gödel [1931], Moravec [1990]*

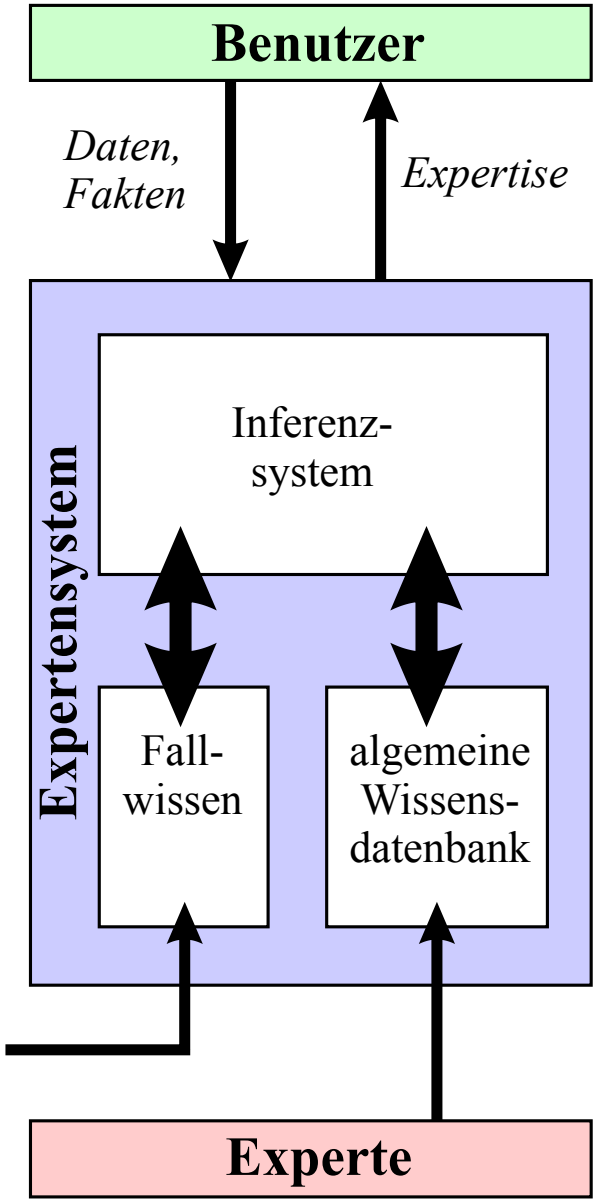
- Gödel zeigt, dass man in jedem genügend mächtigen mathematisch-logischem System (z.B. Peano'sches Axiomensystem zum Aufbau der Arithmetik) Sätze fomulieren kann (Gödelsatz), die einen gültigen Satz des Systems darstellen, in diesem aber nicht beweisbar sind:
  - (selbstbezüglicher) Gödelsatz: „Dieser Satz ist aus den Axiomen nicht ableitbar."

  $\rightarrow$ Annahme: Satz ist ableitbar aus den Axiomen $\Rightarrow$ Satz ist wahr, besagt aber das Ggenteil $\Rightarrow$ Axiomensystem wäre nciht Widerspruchsfrei

  $\leftarrow$ Annahme: Satz ist nicht ableitbar aus den Axiomen $\Rightarrow$ ein wahrer Satz kann nicht aus den Axiomen abgeleitet werden $\Rightarrow$ Axiomensystem ist unvollständig
- Gödels Unvollständigkeitssatz lässt also daran zweifeln, ob Computer wirklich alles berechnen können, was ein Mensch kann, da es Sätze gibt, deren Wahrheit ein Mensch weiß, die aber nicht beweisbar sind (ein Computer ist allerdings auf letztere beschränkt, kann also nie die Wahrheit eines solchen Satze erkennen!)

KI-Zwischenspiel: Expertensysteme (1980-1987)

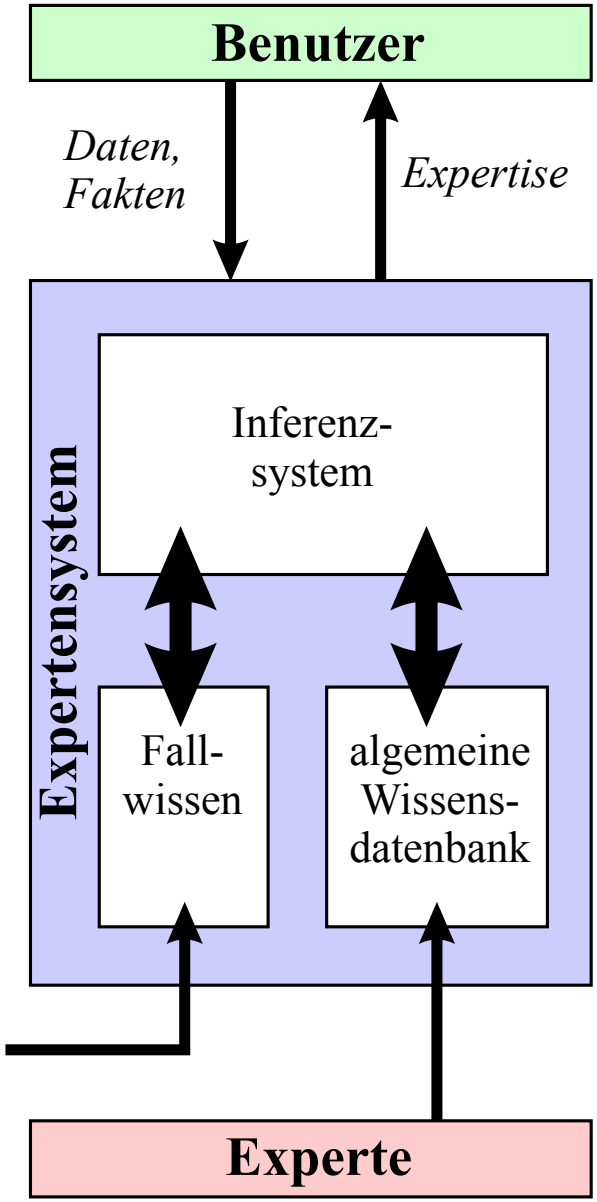
- Wissensbasiertes System, das den Entscheidungsfindungsprozess eines Menschen nachbildet.
- Weiterentwicklung der frühen, Regel-basierten KI-Systeme
- von „Experten“ hand-kodierte Information (Regeln, Wenn-Dann-Beziehungen, ...) ermöglichen es automatisch Schlüsse/Vorhersagen/Pläne/Interpretationen aus gegebenen Situationen/Daten zu ziehen.
- Expertensysteme finden ab 1980 weite Verbreitung (Industrie, Medizin, ...)
- ABER:
  - schwer zu erstellen/zu warten/zu betreiben
  - nicht sehr robust (z.B. bei ungenauen Daten)



Ref: *Ertel [2016]*

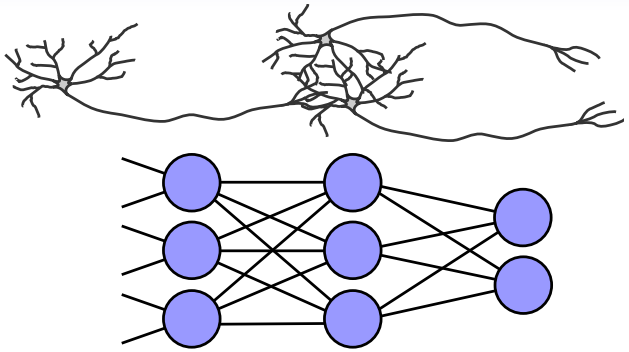
KI-Zwischenspiel: Expertensysteme (1980-1987)

- Wissensbasiertes System, das den Entscheidungsfindungsprozess eines Menschen nachbildet.
- Weiterentwicklung der frühen, Regel-basierten KI-Systeme
- von „Experten“ hand-kodierte Information (Regeln, Wenn-Dann-Beziehungen, ...) ermöglichen es automatisch Schlüsse/Vorhersagen/Pläne/Interpretationen aus gegebenen Situationen/Daten zu ziehen.
- Expertensysteme finden ab 1980 weite Verbreitung (Industrie, Medizin, ...)
- ABER:
  - schwer zu erstellen/zu warten/zu betreiben
  - nicht sehr robust (z.B. bei ungenauen Daten)



Ref: *Ertel [2016]*

Starke KI: Neue Ansätze ab 1990



- **Subsymbolische Systeme:**
  Weiterentwicklung bei **neuronalen Netzen** (komplexere Strukturen können mehr Probleme lösen) und anderen **statistischen Verfahren zur Klassifikation/Lernen**; genetische Algorithmen.
- **Agentensystem:**
  kleines KI-Programm, das mit der Welt interagiert und eine bestimmte (kleine) Aufgabe löst.
- **Multi-Agentensysteme:**
  viele unabhängige (Software-)Agenten (aber auch Menschen, Organisationen...) werden zu großem System zusammengefasst
- **„Embodiment“ & Robotik:**
  Lernen wie ein Kind/Lernen an der Welt: Lernprozess/Entwicklung kognitiver Fähigkeiten beim Menschen nachbilden, KI kann nur durch Interaktion mit der Welt entstehen Teilaufgaben können von Roboterkonstruktion gelöst werden (z.B. Anordnung von Augen,

⇒ **Bottom-Up-Ansatz:** KI als **emergentes Phänomen** in Hierarchie einfacher Subsysteme (z.B. neuronale Netze, Subsumptionsarchitektur)
Vorteile: robust (weniger Anfällig für Fehler/Rauschen), gute Skalierbarkeit

Ref: *Brooks [1990], Braitenberg [1993], Duffy and Joue [2000], Chrisley [2003]*

Starke KI: Neue Ansätze ab 1990

- **Subsymbolische Systeme:**

  Weiterentwicklung bei **neuronalen Netzen** (komplexere Strukturen können mehr Probleme lösen) und anderen **statistischen Verfahren zur Klassifikation/Lernen**; genetische Algorithmen.

- **Agentensystem:**

  kleines KI-Programm, das mit der Welt interagiert und eine bestimmte (kleine) Aufgabe löst.

- **Multi-Agentensysteme:**

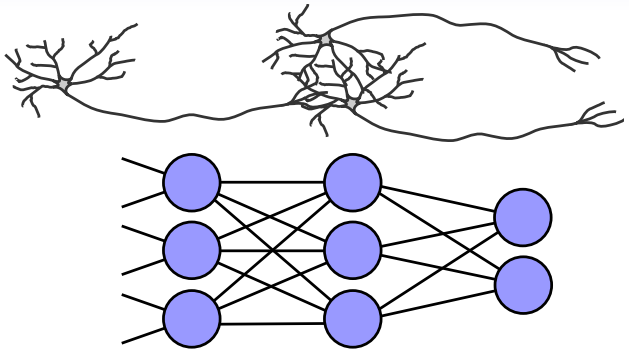
  viele unabhängige (Software-)Agenten (aber auch Menschen, Organisationen...) werden zu großem System zusammengefasst

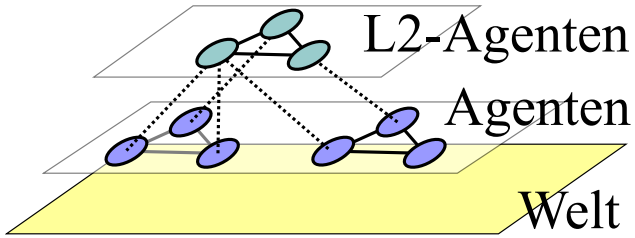
- **„Embodiment“ & Robotik:**

  Lernen wie ein Kind/Lernen an der Welt: Lernprozess/Entwicklung kognitiver Fähigkeiten beim Menschen nachbilden, KI kann nur durch Interaktion mit der Welt entstehen Teilaufgaben können von Roboterkonstruktion gelöst werden (z.B. Anordnung von Augen,

⇒ **Bottom-Up-Ansatz:** KI als **emergentes Phänomen** in Hierarchie einfacher Subsysteme (z.B. neuronale Netze, Subsumptionsarchitektur)

Vorteile: robust (weniger Anfällig für Fehler/Rauschen), gute Skalierbarkeit





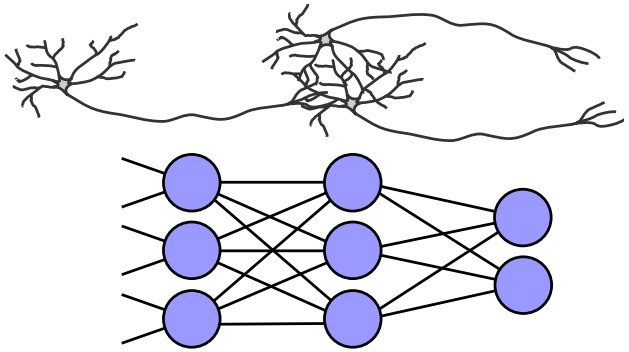
Ref: *Brooks [1990], Braitenberg [1993], Duffy and Joue [2000], Chrisley [2003]*

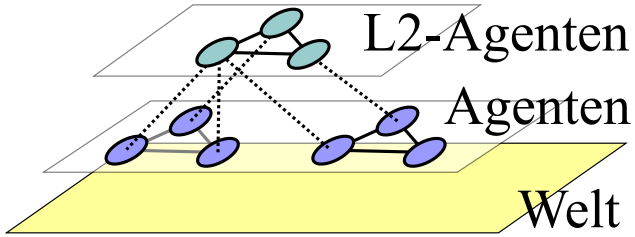
Starke KI: Neue Ansätze ab 1990

- **Subsymbolische Systeme:**
  Weiterentwicklung bei **neuronalen Netzen** (komplexere Strukturen können mehr Probleme lösen) und anderen **statistischen Verfahren zur Klassifikation/Lernen**; genetische Algorithmen.
- **Agentensystem:**
  kleines KI-Programm, das mit der Welt interagiert und eine bestimmte (kleine) Aufgabe löst.
- **Multi-Agentensysteme:**
  viele unabhängige (Software-)Agenten (aber auch Menschen, Organisationen...) werden zu großem System zusammengefasst
- **„Embodiment“ & Robotik:**
  Lernen wie ein Kind/Lernen an der Welt: Lernprozess/Entwicklung kognitiver Fähigkeiten beim Menschen nachbilden, KI kann nur durch Interaktion mit der Welt entstehen Teilaufgaben können von Roboterkonstruktion gelöst werden (z.B. Anordnung von Augen,

⇒ **Bottom-Up-Ansatz:** KI als **emergentes Phänomen** in Hierarchie einfacher Subsysteme (z.B. neuronale Netze, Subsumptionsarchitektur)

Vorteile: robust (weniger Anfällig für Fehler/Rauschen), gute Skalierbarkeit





MISSING IMAGE

Ref: *Brooks [1990], Braitenberg [1993], Duffy and Joue [2000], Chrisley [2003]*

Starke KI: Neue Ansätze ab 1990

- **Subsymbolische Systeme:**

  Weiterentwicklung bei **neuronalen Netzen** (komplexere Strukturen können mehr Probleme lösen) und anderen **statistischen Verfahren zur Klassifikation/Lernen**; genetische Algorithmen.

- **Agentensystem:**

  kleines KI-Programm, das mit der Welt interagiert und eine bestimmte (kleine) Aufgabe löst.

- **Multi-Agentensysteme:**

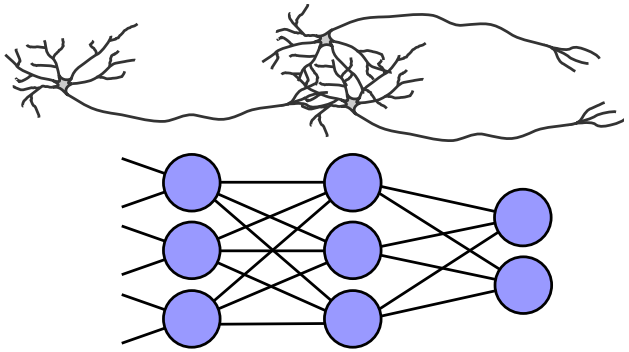
  viele unabhängige (Software-)Agenten (aber auch Menschen, Organisationen...) werden zu großem System zusammengefasst

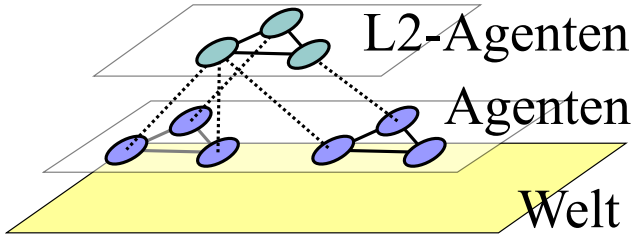
- **„Embodiment“ & Robotik:**

  Lernen wie ein Kind/Lernen an der Welt: Lernprozess/Entwicklung kognitiver Fähigkeiten beim Menschen nachbilden, KI kann nur durch Interaktion mit der Welt entstehen Teilaufgaben können von Roboterkonstruktion gelöst werden (z.B. Anordnung von Augen,

⇒ **Bottom-Up-Ansatz:** KI als **emergentes Phänomen** in Hierarchie einfacher Subsysteme (z.B. neuronale Netze, Subsumptionsarchitektur)

Vorteile: robust (weniger Anfällig für Fehler/Rauschen), gute Skalierbarkeit





MISSING IMAGE

Ref: *Brooks [1990], Braitenberg [1993], Duffy and Joue [2000], Chrisley [2003]*

Starke KI: Stand heute

- Mischung verschiedener Ansätze, keine wirklich neuen Methoden/Ansätze
- sehr populäre: Neuronale Netze („**Deep Learning**“)
- mehr Daten (→ **Big Data**) und mehr Rechenleistung (Moore's Law) ermöglichen bessere Ergebnisse
- viele **vernetzte Systeme** (Smartphones, „Internet of things“, „Smart Cities“, ...)
- Beispiele:
  - 1997: IBM's Schachcomputer DeepBlue gewinnt gegen Garry Kasparov
    wenig intelligenter Ansatz: hohe Rechenleistung, spezialisierte Hardware!
  - 2009: Google baut selbstfahrendes Auto
  - 2011: IBM's KI-System Watson gewinnt bei Jeopardy!
    nutzt Spracherkennung + maschinelles Lernen + große Datenbasis (z.B. Wikipedia) aus strukturierten und unstrukturierten Informationen
  - 2011: Apple/Google/Microsoft veröffentlichen persönliche Assistenten mit Spracherkennung, die einfache Anfragen beantworten können
    Spracherkennung + Wissensbasis auf Servern der Unternehmen
  - 2015: AlphaGo (Google) gewinnt gegen menschlichen Go-Champion:
    nutzt stochastische Suchverfahren + maschinelles Lernen
  - KI-basierte Aktienhandelssysteme (Mustererkennung)

⇒ **Anwendungen, Anwendungen, Anwendungn, ... und wieder Anstrengungen in Richtung starke KI/algemeine KI.**

Starke KI: Stand heute

- Mischung verschiedener Ansätze, keine wirklich neuen Methoden/Ansätze
- sehr populäre: Neuronale Netze („**Deep Learning**“)
- mehr Daten (→ **Big Data**) und mehr Rechenleistung (Moore's Law) ermöglichen bessere Ergebnisse
- viele **vernetzte Systeme** (Smartphones, „Internet of things“, „Smart Cities“, ...)
- Beispiele:
  - 1997: IBM's Schachcomputer DeepBlue gewinnt gegen Garry Kasparov
    wenig intelligenter Ansatz: hohe Rechenleistung, spezialisierte Hardware!
  - 2009: Google baut selbstfahrendes Auto
  - 2011: IBM's KI-System Watson gewinnt bei Jeopardy!
    nutzt Spracherkennung + maschinelles Lernen + große Datenbasis (z.B. Wikipedia) aus strukturierten und unstrukturierten Informationen
  - 2011: Apple/Google/Microsoft veröffentlichen persönliche Assistenten mit Spracherkennung, die einfache Anfragen beantworten können
    Spracherkennung + Wissensbasis auf Servern der Unternehmen
  - 2015: AlphaGo (Google) gewinnt gegen menschlichen Go-Champion:
    nutzt stochastische Suchverfahren + maschinelles Lernen
  - KI-basierte Aktienhandelssysteme (Mustererkennung)

⇒ **Anwendungen, Anwendungen, Anwendungn, ... und wieder Anstrengungen in Richtung starke KI/algemeine KI.**

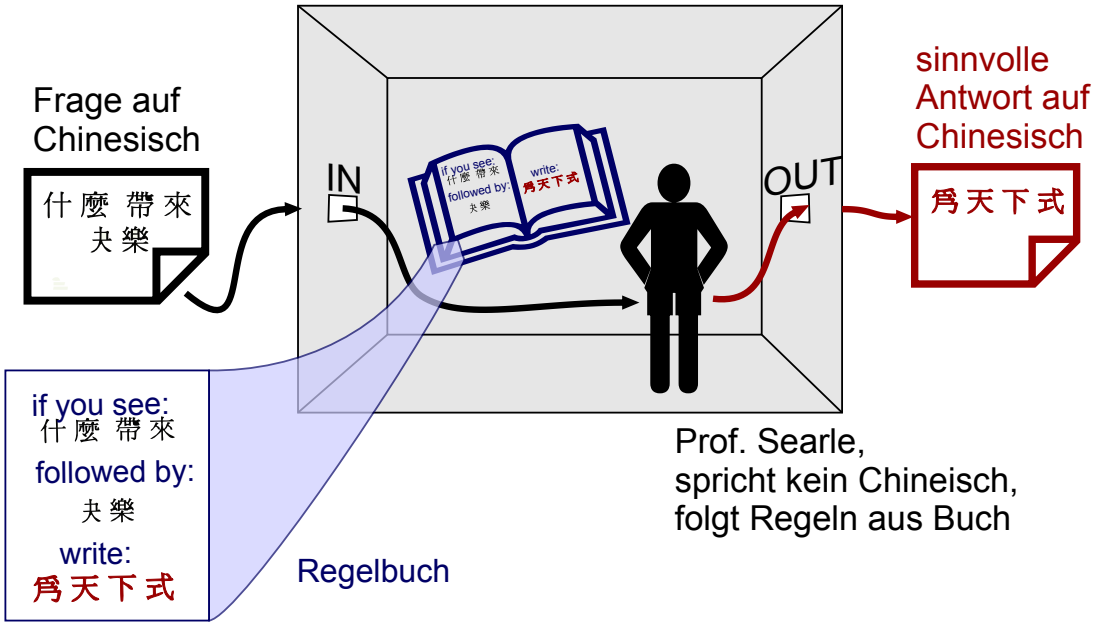
Starke KI: Stand heute

- Mischung verschiedener Ansätze, keine wirklich neuen Methoden/Ansätze
- sehr populäre: Neuronale Netze („**Deep Learning**“)
- mehr Daten (→ **Big Data**) und mehr Rechenleistung (Moore's Law) ermöglichen bessere Ergebnisse
- viele **vernetzte Systeme** (Smartphones, „Internet of things“, „Smart Cities“, ...)
- Beispiele:
  - 1997: IBM's Schachcomputer DeepBlue gewinnt gegen Garry Kasparov
    wenig intelligenter Ansatz: hohe Rechenleistung, spezialisierte Hardware!
  - 2009: Google baut selbstfahrendes Auto
  - 2011: IBM's KI-System Watson gewinnt bei Jeopardy!
    nutzt Spracherkennung + maschinelles Lernen + große Datenbasis (z.B. Wikipedia) aus strukturierten und unstrukturierten Informationen
  - 2011: Apple/Google/Microsoft veröffentlichen persönliche Assistenten mit Spracherkennung, die einfache Anfragen beantworten können
    Spracherkennung + Wissensbasis auf Servern der Unternehmen
  - 2015: AlphaGo (Google) gewinnt gegen menschlichen Go-Champion:
    nutzt stochastische Suchverfahren + maschinelles Lernen
  - KI-basierte Aktienhandelssysteme (Mustererkennung)

⇒ **Anwendungen, Anwendungen, Anwendungn, ... und wieder Anstrengungen in Richtung starke KI/algemeine KI.**

Kritik an starker KI: „Chinese Room“ Argument

Der Philosoph John R. Searle schlägt 1980 Gedankenexperiment vor, um zu zeigen, dass Starke KI unmöglich ist:



- Prof. Searle (=Computer) hat kein Verständnis für Semantik der Fragen/Antworten (=Verstehen), nur für ihre Syntax
- Regelbuch (=Programm) enthält nur Syntax-Regeln (auch keine Semantik)
- Geist/Intelligenz benötigt Semantik/Verstehen

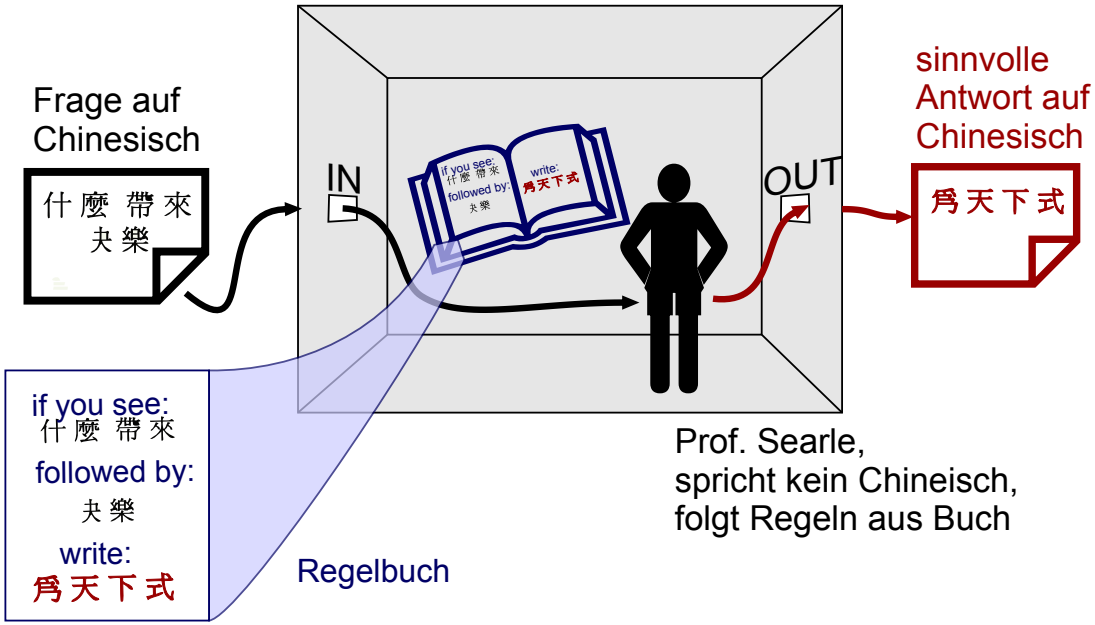
⇒ „Chinese Room“ (und damit KI) zeigen kein echtes Verstehen

- Häufiges Gegenargument: Semantik nicht in Einzelteilen, aber im Gesamtsystem

Ref: *Searle [1980, 1987]*

Kritik an starker KI: „Chinese Room“ Argument

Der Philosoph John R. Searle schlägt 1980 Gedankenexperiment vor, um zu zeigen, dass Starke KI unmöglich ist:



- Prof. Searle (=Computer) hat kein Verständnis für Semantik der Fragen/Antworten (=Verstehen), nur für ihre Syntax
- Regelbuch (=Programm) enthält nur Syntax-Regeln (auch keine Semantik)
- Geist/Intelligenz benötigt Semantik/Verstehen

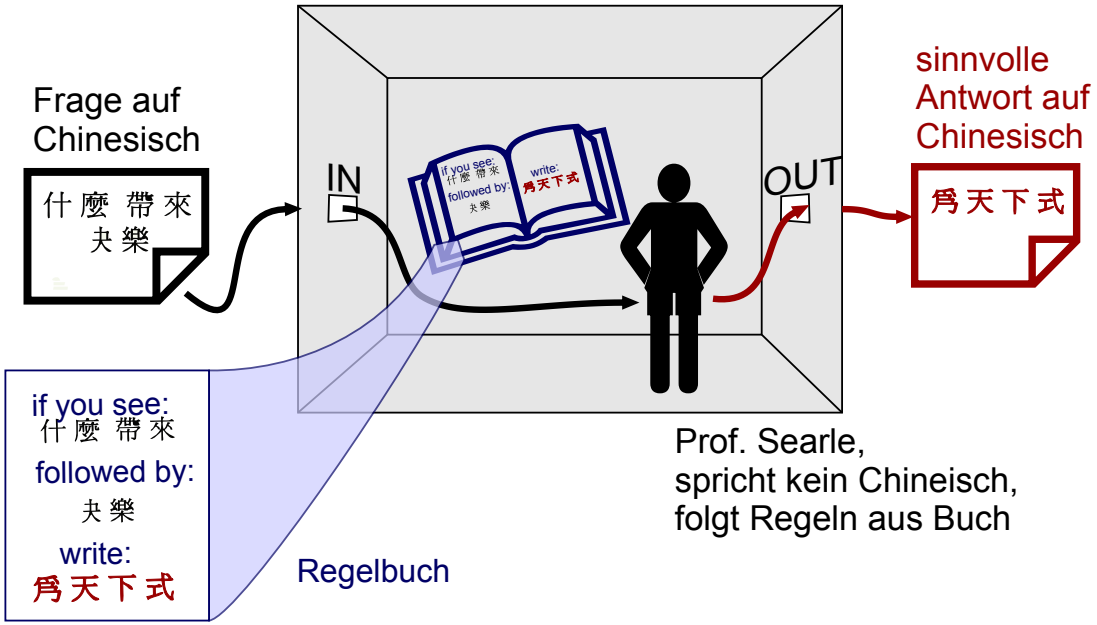
⇒ „Chinese Room“ (und damit KI) zeigen kein echtes Verstehen

- Häufiges Gegenargument: Semantik nicht in Einzelteilen, aber im Gesamtsystem

Ref: *Searle [1980, 1987]*

Kritik an starker KI: „Chinese Room“ Argument

Der Philosoph John R. Searle schlägt 1980 Gedankenexperiment vor, um zu zeigen, dass Starke KI unmöglich ist:



- Prof. Searle (=Computer) hat kein Verständnis für Semantik der Fragen/Antworten (=Verstehen), nur für ihre Syntax
- Regelbuch (=Programm) enthält nur Syntax-Regeln (auch keine Semantik)
- Geist/Intelligenz benötigt Semantik/Verstehen

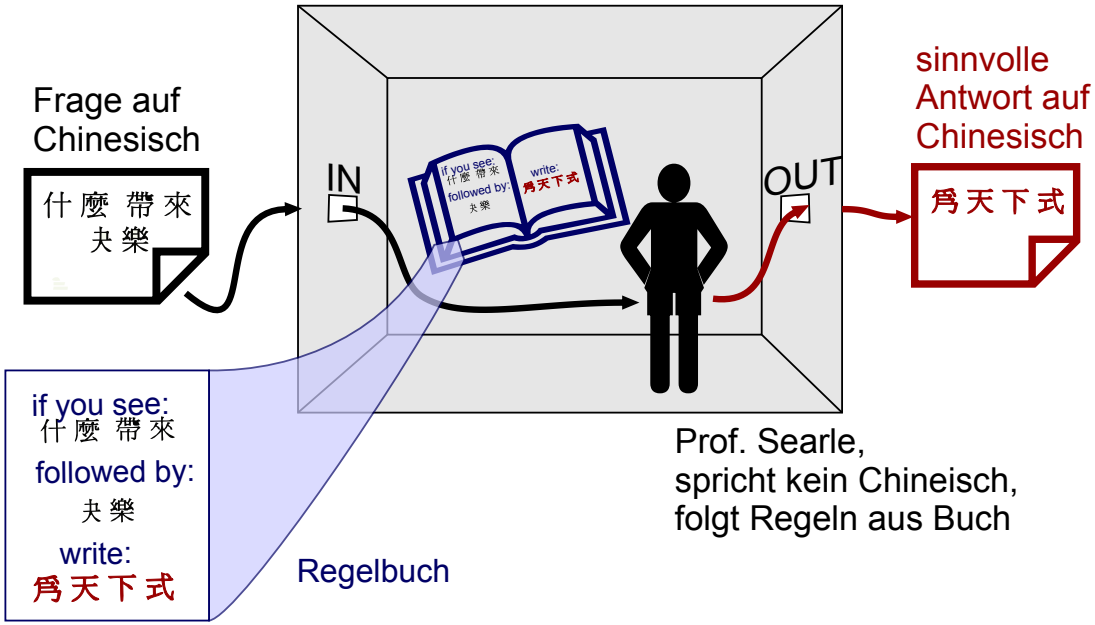
⇒ „Chinese Room“ (und damit KI) zeigen kein echtes Verstehen

- Häufiges Gegenargument: Semantik nicht in Einzelteilen, aber im Gesamtsystem

Ref: *Searle [1980, 1987]*

Kritik an starker KI: „Chinese Room“ Argument

Der Philosoph John R. Searle schlägt 1980 Gedankenexperiment vor, um zu zeigen, dass Starke KI unmöglich ist:



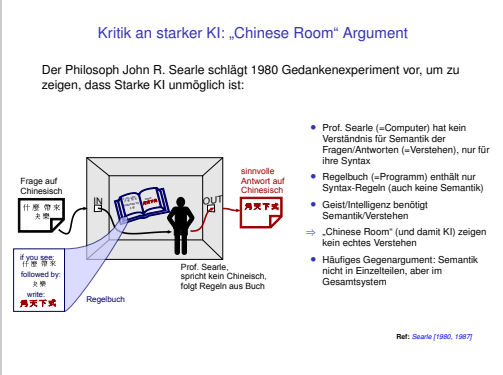
- Prof. Searle (=Computer) hat kein Verständnis für Semantik der Fragen/Antworten (=Verstehen), nur für ihre Syntax
- Regelbuch (=Programm) enthält nur Syntax-Regeln (auch keine Semantik)
- Geist/Intelligenz benötigt Semantik/Verstehen

$\Rightarrow$ „Chinese Room“ (und damit KI) zeigen kein echtes Verstehen

- Häufiges Gegenargument: Semantik nicht in Einzelteilen, aber im Gesamtsystem

Ref: *Searle [1980, 1987]*

Kritik an starker KI: „Chinese Room“ Argument



- Gedankenexperiment:
  1. Es gibt einen Raum, in den beliebige, schriftliche Fragen auf Chinesisch hineingereicht werden können.
  2. Der Raum gibt einen Antwortzettel zurück, auf dem eine plausible und zuverlässig korrekte Antwort steht.
  3. Im Raum befindet sich ein Prof. Searle, der kein Chinesisch spricht, sondern nur Englisch und ein Buch, in dem in englischer Sprache Regeln stehen, mit denen man die Antworten aus den Zeichenfolgen/Bilder der Frage zusammensetzt. Prof. Searle benötigt dazu explizit kein Wissen darüber, was die Zeichen der Ein- und Ausgabe bedeuten.

⇒ Prof. Searle hat kein Verständnis für Fragen/Antworten

⇒ Regelbuch sind „nur“ kontext-freie Regeln, also auch kein Verständnis

⇒ kein echtes Verstehen im Raum vorhanden

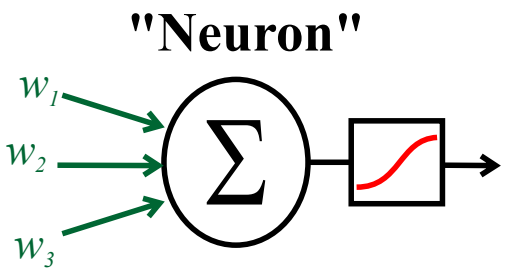
⇒ Der Raum simuliert „nur“ intelligentes Verhalten, ebenso wie KI, d.h. KI zeigt keine echte Intelligenz

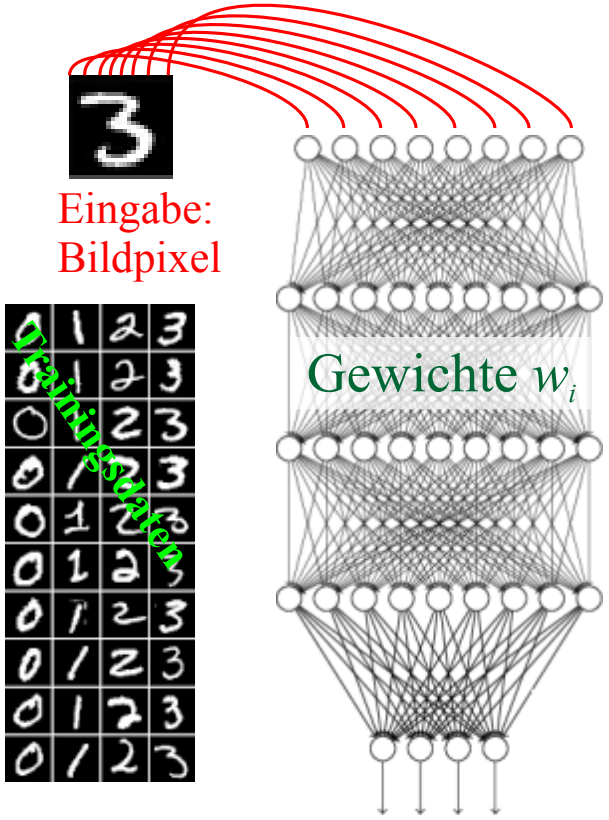
- Searle bezog sich wohl nicht explizit auf Turing-Test, aber implizit
- Gegenargument, z.B.: Gesamtsystem verhält sich Intelligent, obwohl es Einzelteile nicht tun (Buch, Pofessor), ein einzelnes Neuron wird auch nicht als „intelligent“ bezeichnet, das Verhalten des Menschen aber schon

Kritik: Methodische Schwächen am Beispiel neuronale Netze (NN)

Was ist ein NN?

- Neuron: Ausgabe hängt über Schwellwert von gewichteten Eingängen (viele!) ab.
- moderne neuronale Netze: **viele Neuronen, viele Schichten, viele Vernetzungen**
- Anwendung als **Klassifikator** (z.B. Buchstabenerkennung aus Bildern, ...)
- **Lernen:** Präsentiere Beispiele und ändere Synapsen-Gewichte so, dass Klassifizierung möglichst gut
- heute ist spezialisierte Hardware für NN verfügbar (Google „Tensor Processing Unit“, Grafikkarten), die NN sehr schnell ausführen/trainieren kann.





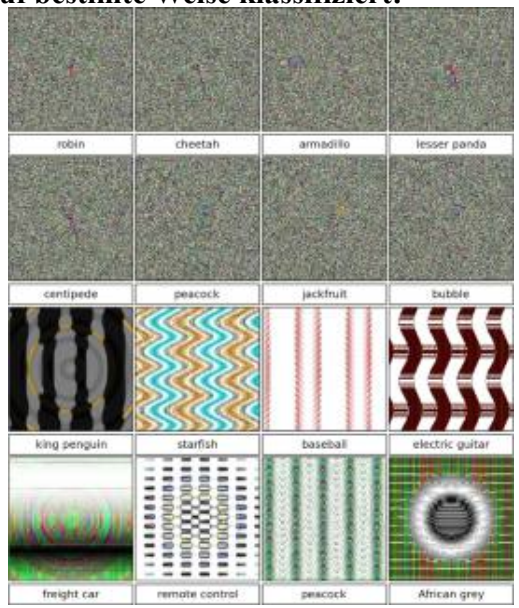
Kritik: Methodische Schwächen am Beispiel neuronale Netze (NN)

Probleme mit NN

- **„Black-box“/mangelnde Überprüfbarkeit:**
  es ist schwer/nicht möglich zu verstehen WARUM eine Eingabe einer bestimmten Klasse zugeordnet wird
- Lernmethoden können in **lokalem Optimum** steckenbleiben (Bergsteigen im Nebel)
- große Mengen an vorklassifizierten **Trainingsdaten nötig** (woher kommen diese? Subjektive Auswahl durch Designer)
- **NN lassen sich leicht täuschen**: NN erkennen Bildstatistik, lassen sich leicht durch zufällige Bilder mit der richtigen Statistik täuschen, also kein echtes Bildverstehen?

unvollständige Trainingsdaten: "Hantel immer mit Hand"

Täuschung: Bilder lernen, die NN auf bestimte Weise klassifiziert:



Ref: *Mordvintsev [2015], Nguyen et al. [2015]*

Kritik: KI, Gesellschaft & Ethik

Welchen Einfluss hat KI-Technologie auf die Gesellschaft?
Wie gehen wir mit KI-Technologien um?

- **Unfälle mit KI-Systemen** (z.B. beim selbstfahrenden Auto):
  - Wer ist bei Unfällen verantwortlich?
  - Wie kann man aus Unfällen lernen?
- **Sicherheit von KI-Systemen**
  - Darf der Nutzer ein KI-System weitertrainieren? Wie kontrolliert man das?
  - Benötigen wir Regeln (ähnlich Ethikregeln in der Medizin) als Schutz? Kann man solche Regeln durchsetzen?
  - allgemein: Wie verhindert man Schäden durch defekte oder böswillige KI-Systeme?
  - Wie verhindert man voreingenommene/diskriminierende KI-Systemen (z.B. „Bias“ durch den Programmierer, Datenauswahl, ...)?
- **Akzeptanz von KI-Entscheidungen:**
  - KI-Systeme gelten als Unvoreingenommen $\Rightarrow$ Entscheidungen sind schwierig anzugreifen
  - Wie kann man KI-Systeme verständlich/sichtbar gestalten? Ihren Einfluss klarmachen? Wie kann man Entscheidungen nachvollziehen?
- **Ändern KI-Systeme die Gesellschaft?**
  - Umgang mit wegfallenden/sich ändernden Jobs/Berufen
  - Einfluss auf den Bildungsstand?

Ref: *O'Neil [2016], Bostrom and Yudkowsky [2014], Bossmann [02016]*

Kritik: KI, Gesellschaft & Ethik

Welchen Einfluss hat KI-Technologie auf die Gesellschaft?
Wie gehen wir mit KI-Technologien um?

- **Unfälle mit KI-Systemen** (z.B. beim selbstfahrenden Auto):
  - Wer ist bei Unfällen verantwortlich?
  - Wie kann man aus Unfällen lernen?
- **Sicherheit von KI-Systemen**
  - Darf der Nutzer ein KI-System weitertrainieren? Wie kontrolliert man das?
  - Benötigen wir Regeln (ähnlich Ethikregeln in der Medizin) als Schutz? Kann man solche Regeln durchsetzen?
  - allgemein: Wie verhindert man Schäden durch defekte oder böswillige KI-Systeme?
  - Wie verhindert man voreingenommene/diskriminierende KI-Systemen (z.B. „Bias“ durch den Programmierer, Datenauswahl, ...)?
- **Akzeptanz von KI-Entscheidungen:**
  - KI-Systeme gelten als Unvoreingenommen $\Rightarrow$ Entscheidungen sind schwierig anzugreifen
  - Wie kann man KI-Systeme verständlich/sichtbar gestalten? Ihren Einfluss klarmachen? Wie kann man Entscheidungen nachvollziehen?
- **Ändern KI-Systeme die Gesellschaft?**
  - Umgang mit wegfallenden/sich ändernden Jobs/Berufen
  - Einfluss auf den Bildungsstand?

Ref: *O'Neil [2016], Bostrom and Yudkowsky [2014], Bossmann [02016]*

Kritik: KI, Gesellschaft & Ethik

Welchen Einfluss hat KI-Technologie auf die Gesellschaft?
Wie gehen wir mit KI-Technologien um?

- **Unfälle mit KI-Systemen** (z.B. beim selbstfahrenden Auto):
  - Wer ist bei Unfällen verantwortlich?
  - Wie kann man aus Unfällen lernen?
- **Sicherheit von KI-Systemen**
  - Darf der Nutzer ein KI-System weitertrainieren? Wie kontrolliert man das?
  - Benötigen wir Regeln (ähnlich Ethikregeln in der Medizin) als Schutz? Kann man solche Regeln durchsetzen?
  - allgemein: Wie verhindert man Schäden durch defekte oder böswillige KI-Systeme?
  - Wie verhindert man voreingenommene/diskriminierende KI-Systemen (z.B. „Bias“ durch den Programmierer, Datenauswahl, ...)?
- **Akzeptanz von KI-Entscheidungen:**
  - KI-Systeme gelten als Unvoreingenommen $\Rightarrow$ Entscheidungen sind schwierig anzugreifen
  - Wie kann man KI-Systeme verständlich/sichtbar gestalten? Ihren Einfluss klarmachen? Wie kann man Entscheidungen nachvollziehen?
- **Ändern KI-Systeme die Gesellschaft?**
  - Umgang mit wegfallenden/sich ändernden Jobs/Berufen
  - Einfluss auf den Bildungsstand?

Ref: *O'Neil [2016], Bostrom and Yudkowsky [2014], Bossmann [02016]*

Kritik: KI, Gesellschaft & Ethik

Welchen Einfluss hat KI-Technologie auf die Gesellschaft?
Wie gehen wir mit KI-Technologien um?

- **Unfälle mit KI-Systemen** (z.B. beim selbstfahrenden Auto):
  - Wer ist bei Unfällen verantwortlich?
  - Wie kann man aus Unfällen lernen?
- **Sicherheit von KI-Systemen**
  - Darf der Nutzer ein KI-System weitertrainieren? Wie kontrolliert man das?
  - Benötigen wir Regeln (ähnlich Ethikregeln in der Medizin) als Schutz? Kann man solche Regeln durchsetzen?
  - allgemein: Wie verhindert man Schäden durch defekte oder böswillige KI-Systeme?
  - Wie verhindert man voreingenommene/diskriminierende KI-Systemen (z.B. „Bias“ durch den Programmierer, Datenauswahl, ...)?
- **Akzeptanz von KI-Entscheidungen:**
  - KI-Systeme gelten als Unvoreingenommen $\Rightarrow$ Entscheidungen sind schwierig anzugreifen
  - Wie kann man KI-Systeme verständlich/sichtbar gestalten? Ihren Einfluss klarmachen? Wie kann man Entscheidungen nachvollziehen?
- **Ändern KI-Systeme die Gesellschaft?**
  - Umgang mit wegfallenden/sich ändernden Jobs/Berufen
  - Einfluss auf den Bildungsstand?

Ref: *O'Neil [2016], Bostrom and Yudkowsky [2014], Bossmann [02016]*

Kritik: KI, Gesellschaft & Ethik

Welchen Einfluss hat KI-Technologie auf die Gesellschaft?
Wie gehen wir mit KI-Technologien um?

- **Unfälle mit KI-Systemen** (z.B. beim selbstfahrenden Auto):
  - Wer ist bei Unfällen verantwortlich?
  - Wie kann man aus Unfällen lernen?
- **Sicherheit von KI-Systemen**
  - Darf der Nutzer ein KI-System weitertrainieren? Wie kontrolliert man das?
  - Benötigen wir Regeln (ähnlich Ethikregeln in der Medizin) als Schutz? Kann man solche Regeln durchsetzen?
  - allgemein: Wie verhindert man Schäden durch defekte oder böswillige KI-Systeme?
  - Wie verhindert man voreingenommene/diskriminierende KI-Systemen (z.B. „Bias“ durch den Programmierer, Datenauswahl, ...)?
- **Akzeptanz von KI-Entscheidungen:**
  - KI-Systeme gelten als Unvoreingenommen $\Rightarrow$ Entscheidungen sind schwierig anzugreifen
  - Wie kann man KI-Systeme verständlich/sichtbar gestalten? Ihren Einfluss klarmachen? Wie kann man Entscheidungen nachvollziehen?
- **Ändern KI-Systeme die Gesellschaft?**
  - Umgang mit wegfallenden/sich ändernden Jobs/Berufen
  - Einfluss auf den Bildungsstand?

Ref: *O'Neil [2016], Bostrom and Yudkowsky [2014], Bossmann [02016]*

Kritik: KI, Gesellschaft & Ethik

Welchen Einfluss hat KI-Technologie auf die Gesellschaft?
Wie gehen wir mit KI-Technologien um?

- **Unfälle mit KI-Systemen** (z.B. beim selbstfahrenden Auto):
  - Wer ist bei Unfällen verantwortlich?
  - Wie kann man aus Unfällen lernen?
- **Sicherheit von KI-Systemen**
  - Darf der Nutzer ein KI-System weitertrainieren? Wie kontrolliert man das?
  - Benötigen wir Regeln (ähnlich Ethikregeln in der Medizin) als Schutz? Kann man solche Regeln durchsetzen?
  - allgemein: Wie verhindert man Schäden durch defekte oder böswillige KI-Systeme?
  - Wie verhindert man voreingenommene/diskriminierende KI-Systemen (z.B. „Bias" durch den Programmierer, Datenauswahl, ...)?
- **Akzeptanz von KI-Entscheidungen:**
  - KI-Systeme gelten als Unvoreingenommen ⇒ Entscheidungen sind schwierig anzugreifen
  - Wie kann man KI-Systeme verständlich/sichtbar gestalten? Ihren Einfluss klarmachen? Wie kann man Entscheidungen nachvollziehen?
- **Ändern KI-Systeme die Gesellschaft?**
  - Umgang mit wegfallenden/sich ändernden Jobs/Berufen
  - Einfluss auf den Bildungsstand?

Ref: *O'Neil [2016], Bostrom and Yudkowsky [2014], Bossmann [02016]*

> ***Asimov's Robotergesetzte:***
> ***0.*** *Roboter darf der Menschheit nicht schaden.*
> ***1.*** *Roboter darf keinem Menschen schaden zufügen, oder dies durch nichts-tun zulassen.*
> ***2.*** *Roboter muss den Befehlen eines Menschen folgen, außer sie widersprechen Regel 1.*
> ***3.*** *Roboter muss sich selbst schützen, außer dies widerspricht Regel 1 oder 2.*
>
> — ISAAC ASIMOV, 1942

Literatur

Julia Bossmann. Top 9 ethical issues in artificial intelligence | world economic forum, 02016. URL https://www.weforum.org/agenda/2016/10/top-10-ethical-issues-in-artificial-intelligence/. accessed 2017-07-19.

Nick Bostrom and Eliezer Yudkowsky. The ethics of artificial intelligence. In *The Cambridge handbook of artificial intelligence*, pages 316–334. Cambridge University Press, Cambridge, 2014. URL https://intelligence.org/files/EthicsofAI.pdf.

Valentin Braitenberg. *Vehikel. Experimente mit kybernetischen Wesen*. Rohwolt Taschenbuchverlag, 1993. ISBN 978-3499195310.

Rodney A. Brooks. Elephants don't play chess. *Robotics and Autonomous Systems*, 6(1-2):3–15, 1990. doi:10.1016/S0921-8890(05)80025-9.

Ron Chrisley. Embodied artificial intelligence. *Artificial Intelligence*, 149(1):131–150, 2003. doi:10.1016/S0004-3702(03)00055-9.

Brian R Duffy and Gina Joue. Intelligent robots: The question of embodiment. In *Proc. of the Brain-Machine Workshop*, 2000.

Wolfgang Ertel. *Grundkurs Künstliche Intelligenz*. Springer Fachmedien Wiesbaden, Wiesbaden, 2016. ISBN 978-3-658-13548-5. doi:10.1007/978-3-658-13549-2.

Howard Gardner. *Frames of Mind: The Theory of Multiple Intelligences*. Basic Books, 3 edition, 2011. ISBN 9780465024339.

Kurt Gödel. Über formal unentscheidbare sätze der principia mathematica und verwandter systeme i. *Monatshefte für Mathematik und Physik*, 38-38(1):173–198, 1931. doi:10.1007/BF01700692.

Warren S. McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5(4):115–133, 1943. doi:10.1007/BF02478259.

Hans Moravec. *Mind Children: The Future of Robot and Human Intelligence*. Harvard University Press, reprint edition, 1990. ISBN 9780674576186.

Alexander Mordvintsev. Research blog: Inceptionism: Going deeper into neural networks, 2015. URL https://research.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html. accessed 2017-07-18.

Allen Newell and Herbert A. Simon. Computer science as empirical inquiry: symbols and search. *Communications of the ACM*, 19(3):113–126, 1976. doi:10.1145/360018.360022.

Allen Newell, John C. Shaw, and Herbert A. Simon. Report on a general problem-solving program. In *Proceedings of the International Conference on Information Processing*, pages 256–264, 1959.

Anh Nguyen, Jason Yosinski, and Jeff Clune. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

Cathy O'Neil. *Weapons of Math Destruction*. Crown Publishing Group, 2016. ISBN 9780553418811.

Frank Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386–408, 1958.

John Searle. Minds and brains without programs. *Mindwaves*, pages 209–233, 1987.

John R. Searle. Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3):417–457, 1980. doi:10.1017/S0140525X00005756.

Claude E. Shannon. Programming a computer for playing chess. *Philosophical Magazine*, 41 (Ser. 7)(314), 1950.

C. Spearman. "general intelligence,"objectively determined and measured. *The American Journal of Psychology*, 15(2):201, 1904. doi:10.2307/1412107.

A. M. Turing. I. – COMPUTING MACHINERY AND INTELLIGENCE. *Mind*, LIX(236):433–460, 1950. doi:10.1093/mind/LIX.236.433.

J. von Neumann. Zur theorie der gesellschaftsspiele. *Mathematische Annalen*, 100(1):295–320, 1928. doi:10.1007/BF01448847.

J. von Neumann. First draft of a report on the EDVAC. Technical report, Moore School of Electrical Engineering, University of Pennsylvania, 1945.

Joseph Weizenbaum. ELIZA – a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1):36–45, 1966.

Alfred North Whitehead and Bertrand Russel. *Principia Mathematica*, volume 1. Cambridge University Press, 2 edition, 1963.