

Künstliche Intelligenz, was ist das? Wissensverarbeitung und Handlungsplanung in Kognitiven Systemen

Prof. Dr. Birte Glimm | 27.04.2022



Gliederung

- Begriffseinordnung
- Wissensrepräsentation und automatisches Schlussfolgern
- Agenten und Aktionen
- Automatisiertes Planen

Kognitive Systeme

- Der Begriff „Kognition“ bezieht sich traditionell auf die menschliche Fähigkeit Informationen zu erlangen und zu verarbeiten
 - z.B. etwas (wieder) zu **erkennen**
- Ein System das diesen Prozess imitiert muss über verschiedene Fähigkeiten verfügen:
 - **Wahrnehmung** – Daten mittels Sensoren erfassen
 - **Gedächtnis** – Erfahrungen speichern
 - **Lernen** – Ähnliches und Unterschiedliches erkennen
 - **Schlussfolgern** – Wissen ableiten und Entscheidungen treffen
 - **Handeln** – Basierend auf den Entscheidungen handeln
- Menschliche Wahrnehmung, Gedächtnis und Lernen sind meist **low-level** „unbewusste“ Prozesse
- Schlussfolgern und Handeln brauchen **higher-level** (aktives) „Denken“

Bezug zur Künstlichen Intelligenz

- In der Informatik haben kognitive Systeme einen engen Bezug zur künstlichen Intelligenz
- Was heißt Intelligenz?
 - das Vermögen zu **denken** und rational zu **handeln**
- Rational denken:
 - ableiten von Wissen auf Bekanntem
- Rationales Handeln:
 - bestmöglich ein Ziel erreichen

Wie können wir Intelligenz testen oder erkennen?

- Möglichkeit 1: Festlegung der wichtigsten Fähigkeiten intelligenter Wesen
 - zählen, hämmern, musizieren, malen...
- Problem: Die meisten Fähigkeiten haben Computers bereits – sind Sie daher intelligent?



<https://www.youtube.com/embed/S4PYI6TzqYk>

Wie können wir Intelligenz testen oder erkennen?

- Möglichkeit 2: **Der Turing Test**
 - Ein Mensch kommuniziert mit einem Computer über Textnachrichten
 - Der Computer besteht den Test, wenn der Mensch davon überzeugt ist, mit einem Menschen zu kommunizieren

- 69 Jahre später, ist der Test immer noch relevant



Alan Turing
Jon Callas, San Jose, USA, [CC BY 2.0](#)
via [Wikimedia Commons](#)



I'm not a robot



reCAPTCHA

[Privacy](#) - [Terms](#)

Der Turing Test

- Welche Fähigkeiten braucht ein Computer, um den Test zu bestehen?
 - **Verarbeitung natürlicher Sprache**, um zu kommunizieren
 - **Wissensrepräsentation**, um Informationen zu speichern
 - **Automatisches Schlussfolgern**, um Wissen abzuleiten
 - **Maschinelles Lernen**, um sich auf neue Situation anzupassen
 - **Psychologische Modelle**, um wie ein Mensch zu interagieren

Kognitionswissenschaften

- Den Turing Test bestehen – das ultimative Ziel für die KI?
 - Wenn Flugzeugbauer das Ziel hätten Maschinen zu bauen, die wie Tauben aussehen und die Tauben täuschen könnten, hätten wir dann Hubschrauber?
- Eher ist das Ziel, **die der Intelligenz zu Grunde liegenden Prinzipien zu verstehen**
- Dies wird im breiten Feld der **Kognitionswissenschaften** durchgeführt
- ein multidisziplinäres Feld, das Biologie, Psychologie, Mathematik, Linguistik, Wirtschaftswissenschaften, und Informatik umfasst

Wissensrepräsentation, Schlussfolgern und Planen

- Um KI Systeme zu realisieren, müssen **Schlussfolgerungs-** und Interaktionsmodelle entwickelt werden
- Mit einfachen formal-logischen Modellen kann man bereits KI Systeme entwickeln, die Rätsel lösen, Spiele spielen, Navigationshilfe geben
- **Menschliches** Verhalten ist allerdings deutlich komplexer
 - d.h. Menschen wenden zum Problemlösen nicht immer **formale Logik** oder **Algorithmen** an



Gliederung

- Begriffseinordnung
- **Wissensrepräsentation und automatisches Schlussfolgern**
- Agenten und Aktionen
- Automatisiertes Planen

Wissensrepräsentation

- Was ist Wissen?
 - Generelle Informationen über die Welt (z.B. Wikipedia/Wikidata)
 - Bsp.: „Die Stadt Ulm befindet sich in Baden-Württemberg“
- Um ein Problem zu lösen, muss man nicht alles wissen; meist reicht **domänen-spezifisches** Wissen
 - Bsp.: „Die Buslinie 5 startet an der Uni Süd“ (Verkehrsdomäne in Ulm)
- Was ist eine **Repräsentation**?
 - Die Art wie Wissen ausgedrückt ist (z.B. Text, Tabelle, Bild oder Formel)
- Wissen ist mehr als nur Daten!
 - Ein jpeg Bild ist mehr als eine Abfolge von Bildpunkten
- Was ist also der Unterschied?

Semantik

- Semantik ist ein genereller Begriff, um **objekthinärente Eigenschaften** zu beschreiben; im Gegensatz zur Darstellung oder Form des Objekts
- Einige Dinge können unterschiedlich dargestellt werden:
 - Verschiedene Bildformate (jpeg, png, gif, . . .) für dasselbe Bild
 - Die Zahl 5 kann als „fünf“ oder „101“ oder als „(2 + 3)“ geschrieben werden
 - Chinesische und Deutsche Sätze könne dasselbe bedeuten, obwohl sie komplett unterschiedlich aussehen
- Semantik ist das was bleibt, wenn man von den syntaktischen Unterschieden (in der Form) absieht
- Einige Repräsentation sind allerdings einfacher zu verarbeiten
 - z.B. sind Binärzahlen für einen Computer einfacher zu verarbeiten als Dezimalzahlen

Logik

- Seit dem Altertum versuchen Philosophen „**Denkregeln**“ zu formulieren
- Nennenswerte Beispiele solcher Regeln sind **logische Syllogismen** wie:

Sokrates ist ein Mensch
Alle Menschen sind sterblich

∴ Sokrates ist sterblich

- Salopp gesprochen definiert Logik Prinzipien, wie neues Wissen als **Konsequenz** aus vorhandenem Wissen abgeleitet werden kann

Automatisches Schlussfolgern

- Schlussfolgern ist der Prozess logische Konsequenzen aus bestehendem Wissen abzuleiten
- Oft gibt es **unendlich** viele logische Konsequenzen
 - unmöglich alle abzuleiten
- In der Praxis reicht es zumeist zu **prüfen**, ob eine gegebene Aussage eine logische Konsequenz ist
- Ist es möglich ein **Computerprogramm** zu schreiben, was diese Aufgabe **automatisiert**?

Automatisches Schlussfolgern

- Hier ein Programm das prüft ob $a - b = 2$ aus $a + b = 4$ und $a \cdot b = 3$ folgt (folgt es?):
 1. Wähle beliebige Werte für a und b
 2. Prüfe ob $a + b = 4$ und $a \cdot b = 3$ für diese Werte gilt
 3. Wenn nicht, gehe zu Schritt 1 (um neue Werte zu wählen)
 4. Prüfe ob $a - b = 2$ für die gewählten Werte gilt
 5. Wenn ja, antworte „folgt“, sonst antworte „folgt nicht“
- Gibt es Probleme mit diesem Programm?
 - Kann das Programm fälschlicherweise ‘folgt’ antworten?
 - **ja**, z.B. wenn $a = 3$ und $b = 1$ in Schritt 1 gewählt werden
 - Kann das Programm fälschlicherweise „folgt nicht“ antworten?
 - **nein**, wenn es „folgt nicht“ antwortet, haben wir Werte, so dass $a + b = 4$ und $a \cdot b = 3$, aber $a - b \neq 2$
 - Terminiert das Programm immer?
 - **nein**, es kann in einer Dauerschleife zwischen den Schritten 1–3 laufen, indem falsche Werte gewählt werden

Algorithmen zum Automatischen Schlussfolgern

- Es gibt diverse Algorithmen zum automatischen Schlussfolgern
- Typische Algorithmen sind genereller als das eben gezeigte Spezialprogramm
- **Aber:** Nicht jeder Algorithmus ist nützlich (wie eben gesehen)
- Was sind **sinnvolle** Eigenschaften von Algorithmen?
 - **Korrektheit:** jede abgeleitete Konsequenz folgt tatsächlich logisch den gegebenen Eingabeaussagen
 - **Vollständigkeit:** der Algorithmus kann alle logischen Konsequenzen ableiten
 - **Terminierung:** für alle Eingaben gibt der Algorithmus nach endlicher Zeit eine Antwort (folgt/folgt nicht)
 - **Geschwindigkeit:** die Antworten kommen relativ schnell



Gliederung

- Begriffseinordnung
- Wissensrepräsentation und automatisches Schlussfolgern
- **Agenten und Aktionen**
- Automatisiertes Planen

Agenten und Aktionen

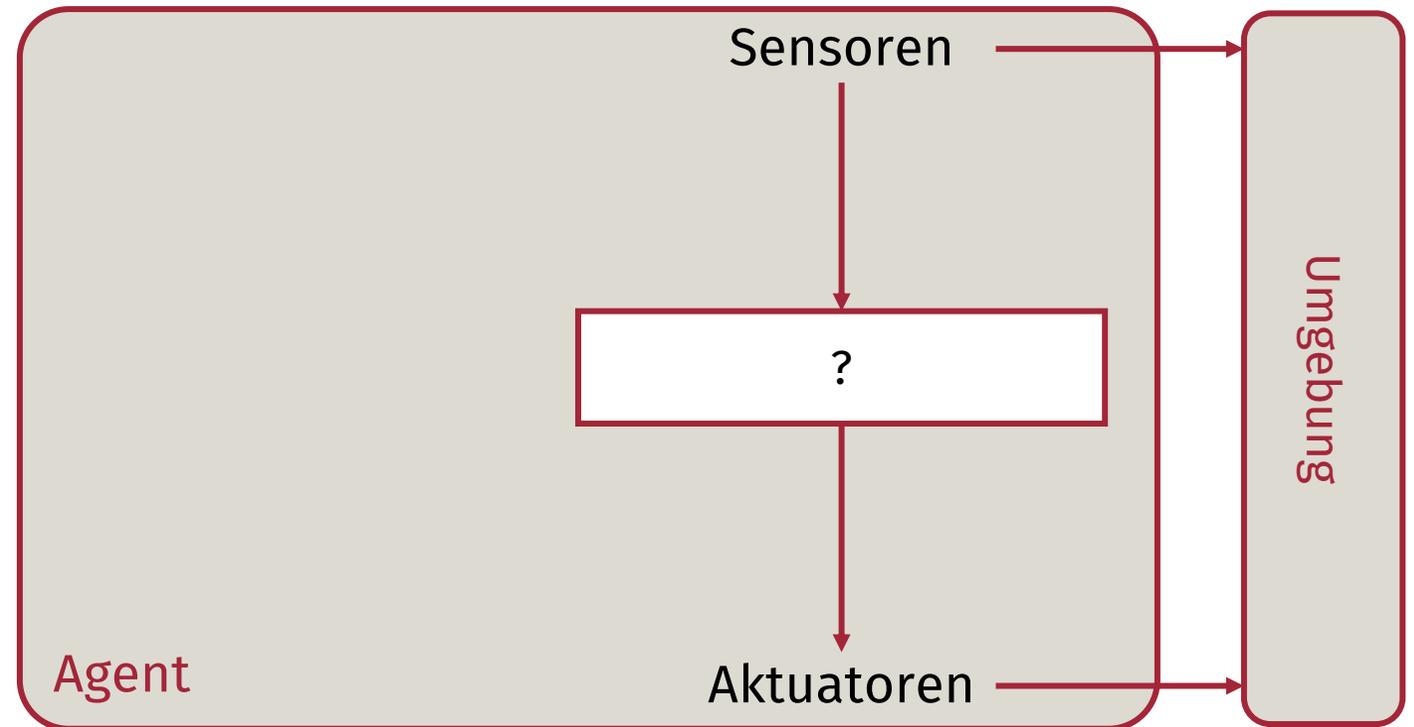
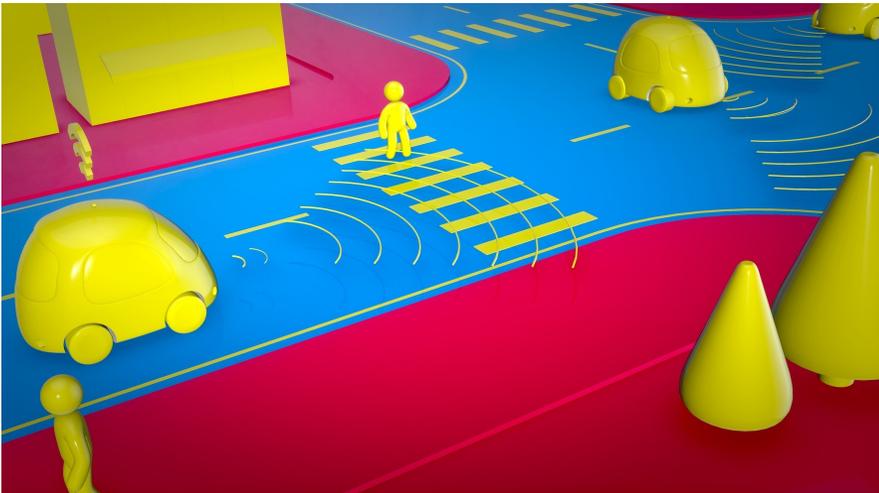
- Es gibt bereits zahlreiche, umfangreiche, formale Wissensbasen (z.B. Wikidata mit 97 Mio. Entitäten, Stand 20.04.2022)
- Um eine KI zu entwickeln, reicht es nicht, schnell logische Konsequenzen abzuleiten
- Jemand muss das Programm nutzen, um mit der „Welt zu interagieren“: so ein Objekt wird **Agent** genannt
- Ein Agent kann eine physische Entität (z.B. ein Staubsaugroboter) oder auch ein Computerprogramm sein



Hersson Piratoba, [CC BY-NC-ND 2.0](#)
via Flickr

Agenten und Aktionen

- Abstrakt gesprochen, kann ein Agent alles sein, was die **Umgebung** mittels **Sensoren** wahrnimmt und durch **Aktuatoren** mit der Umgebung agiert



Agenten und Aktionen

- Beispiele einfacher (elektro-mechanischer) Agenten:



© Raimond Spekking, CC BY-SA 4.0
via Wikimedia Commons

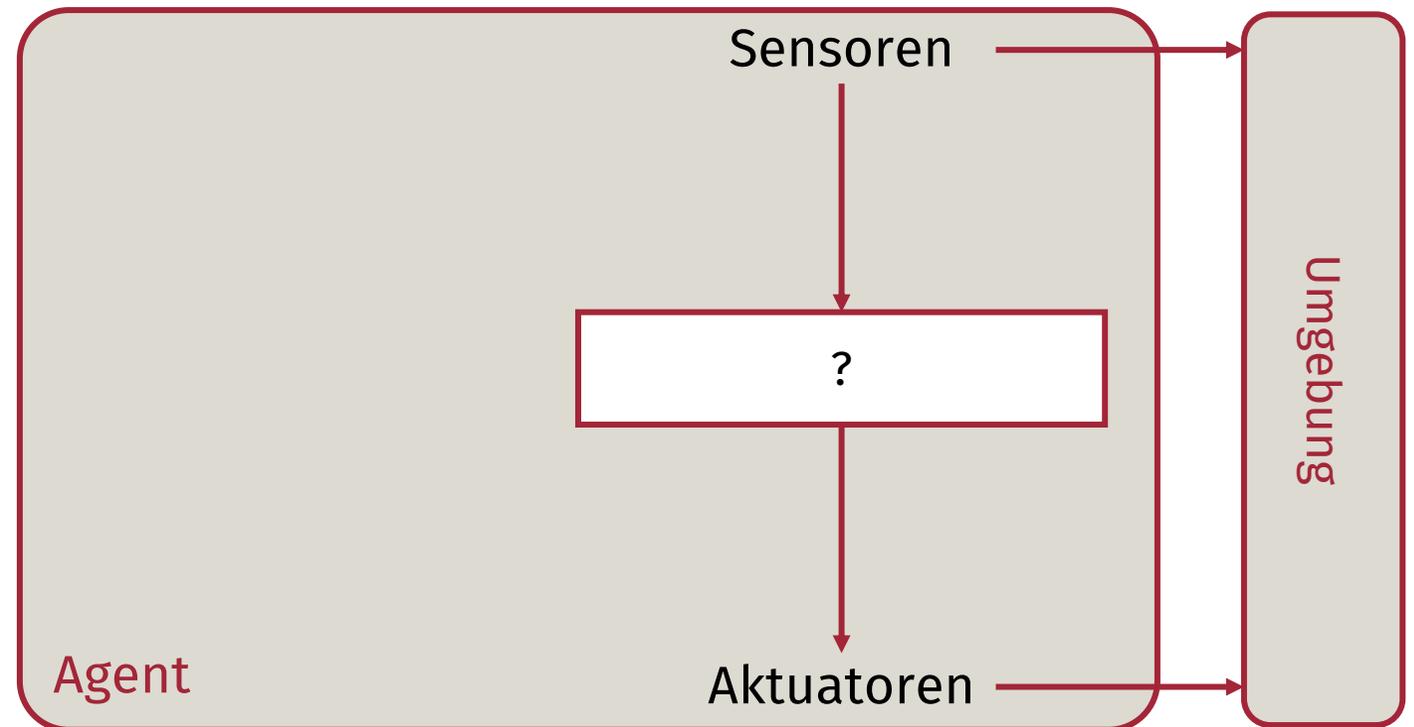
- Diese Agenten erfüllen Ihren Zweck, aber sind sie intelligent?
- Wie kann man entscheiden, ob ein Agent intelligenter ist als ein anderer?
- Um diese Fragen zu beantworten, müssen wir formale Modelle von Agenten, Aktionen und der Umgebung betrachten

Änderungen der Umgebung

- Wie können wir Änderungen in der Umgebung abbilden?
- Die Umgebung kann sich durch Ausführung von **Aktionen** ändern
 - Aktion „gehe von a nach b“ kann den Ort einer Spielfigur ändern
- Die Umgebung kann sich auch ohne Ausführung einer Aktion ändern
 - z.B. durch Handlungen anderer Agenten
- Mit einer Welt zu arbeiten, die sich **willkürlich** ändert, ist sehr herausfordernd für die Entwicklung intelligenter Systeme

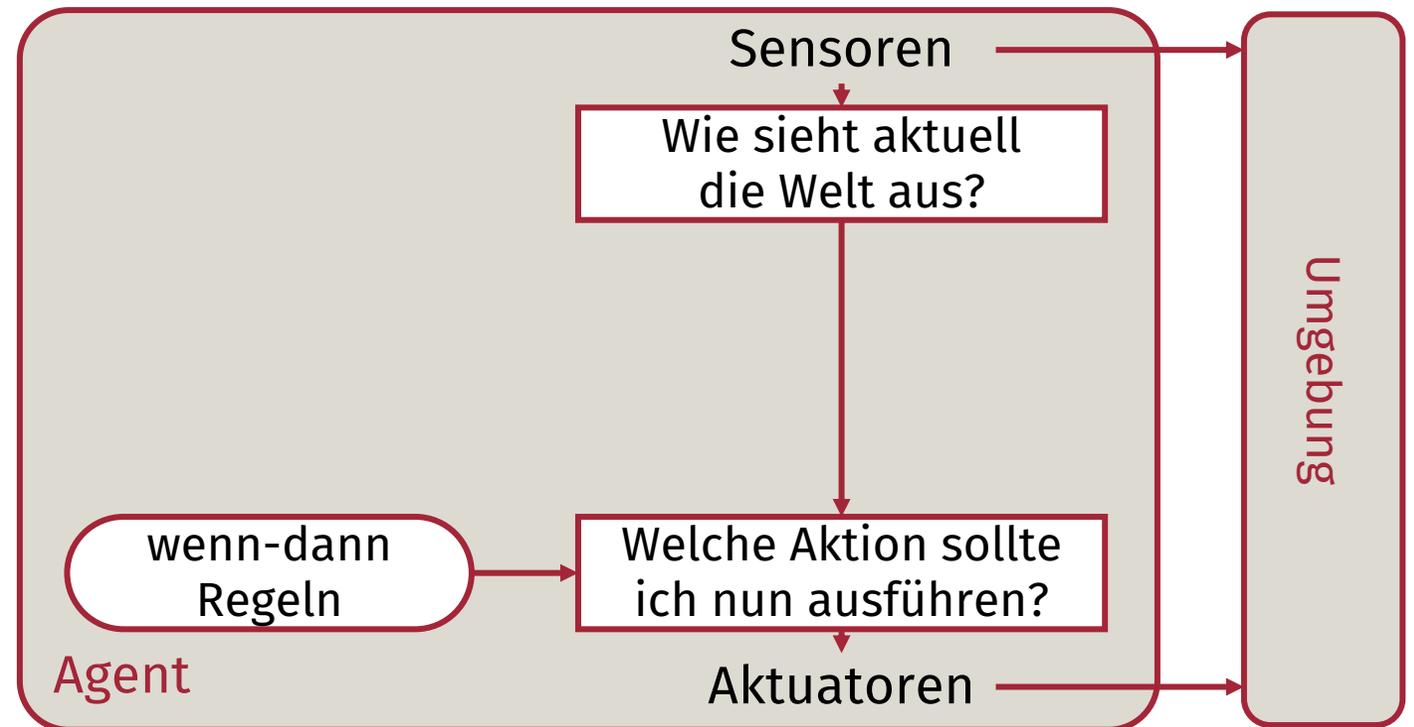
Agenten

- Aktionen werden durch **Agenten** ausgeführt, die:
 - ihre **Umgebung** – den für sie relevanten Teil der Welt – durch **Sensoren** wahrnehmen
 - die **Wahrnehmungen** bilden die **Weltsicht** des Agenten
 - **Aktionen** werden auf der Umgebung durchgeführt



Reflexagenten

- Die einfachsten Agenten sind **Reflexagenten**
 - die Sicht des Agenten auf die Umgebung wird bestimmt durch **die letzte Beobachtung**
 - die Aktionen werden durch einfache „wenn ... , dann ...“ Regeln bestimmt



Reflexagenten

- Beispiel: Thermostat

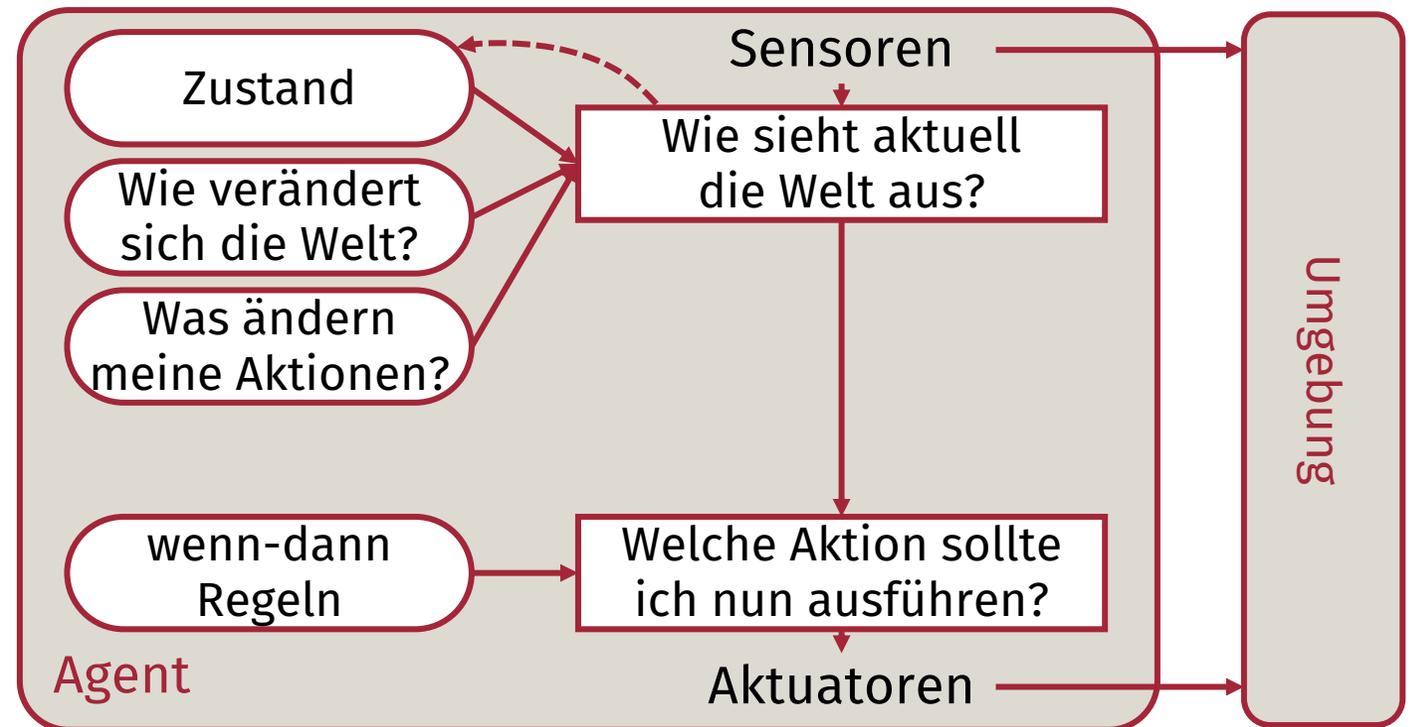
1. Wenn die Außentemperatur unter 20° ist,
dann **schalte die Heizung ein**
2. Wenn die Außentemperatur über 25° ist,
dann **schalte die Heizung aus**



- Der menschliche Reflex – z.B. die Hand von einer heißen Herdplatte wegzuziehen – basiert auf diesem Prinzip
- „Denken“ ist nicht beteiligt

Agenten mit internem Zustand

- Komplexere Agenten müssen ihre Entscheidungen basierend auf **verschiedenen Beobachtungen** treffen
 - ein **interner Zustand** speichert frühere Beobachtungen
 - Aktionen werden durch Änderungen des internen Zustands ausgelöst



Agenten mit internem Zustand

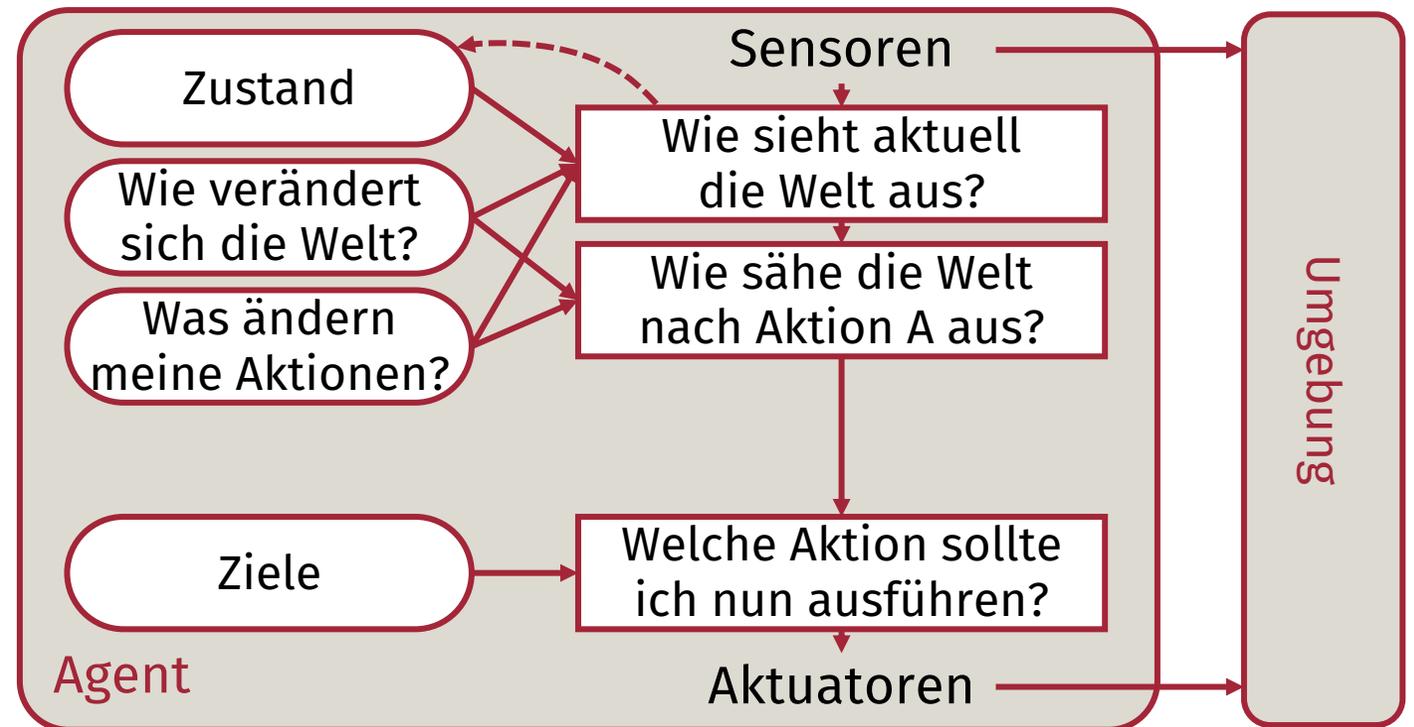
- Beispiel: Lichtschalter mit einem Bewegungsmelder und einer Zeitschaltung
 - Das Ausschalten des Lichts hängt nicht nur von der letzten Beobachtung ab (keine Bewegung), sondern auch von der Beobachtungshistorie (keine Bewegung für eine bestimmte Zeit)
 - Andere Beispiele: Fahrstühle, Kaffeemaschinen, Mikrowellen, Waschmaschinen, ... (alles mit Schaltern)



© Raimond Spekking, [CC BY-SA 4.0](#)
via Wikimedia Commons

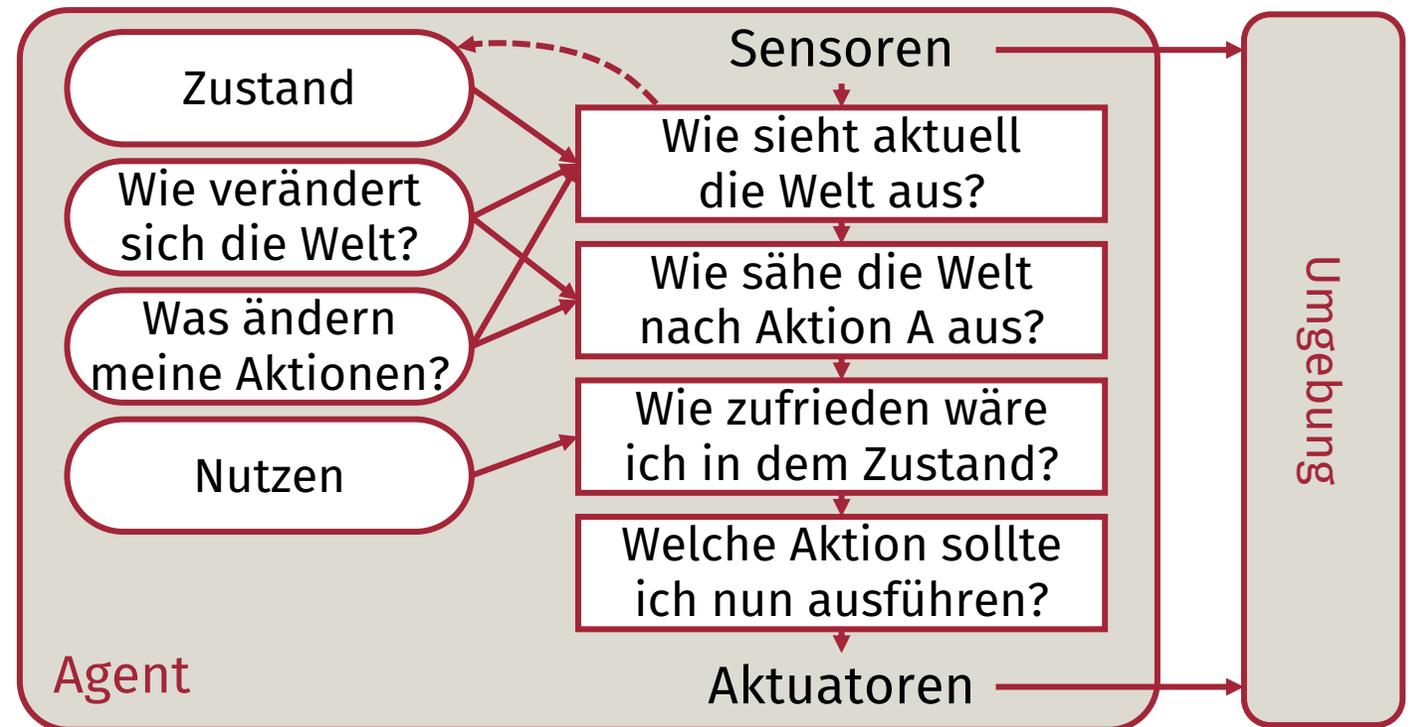
Ziel-basierte Agenten

- Agenten mit internem Zustand sind sehr ausdrucksmächtig
 - komplexe Geräte wie Computer lassen sich damit modellieren
- Ob ein Agent intelligent oder nützlich ist, können wir aber nur entscheiden, wenn wir den Zweck der Handlungen kennen
- Ein **Ziel-basierter Agent** versucht Handlungen durchzuführen, um ein bestimmtes Ziel zu erreichen



Nutzen-basierte Agenten

- Oft gibt es viele Möglichkeiten ein Ziel zu erreichen
- Manchmal können nicht alle Ziele gleichzeitig erreicht werden
- Ein intelligenter Agent würde versuchen, um einen maximalen **Nutzen** zu erreichen
 - ein Maß des durch die Aktionen Erreichten

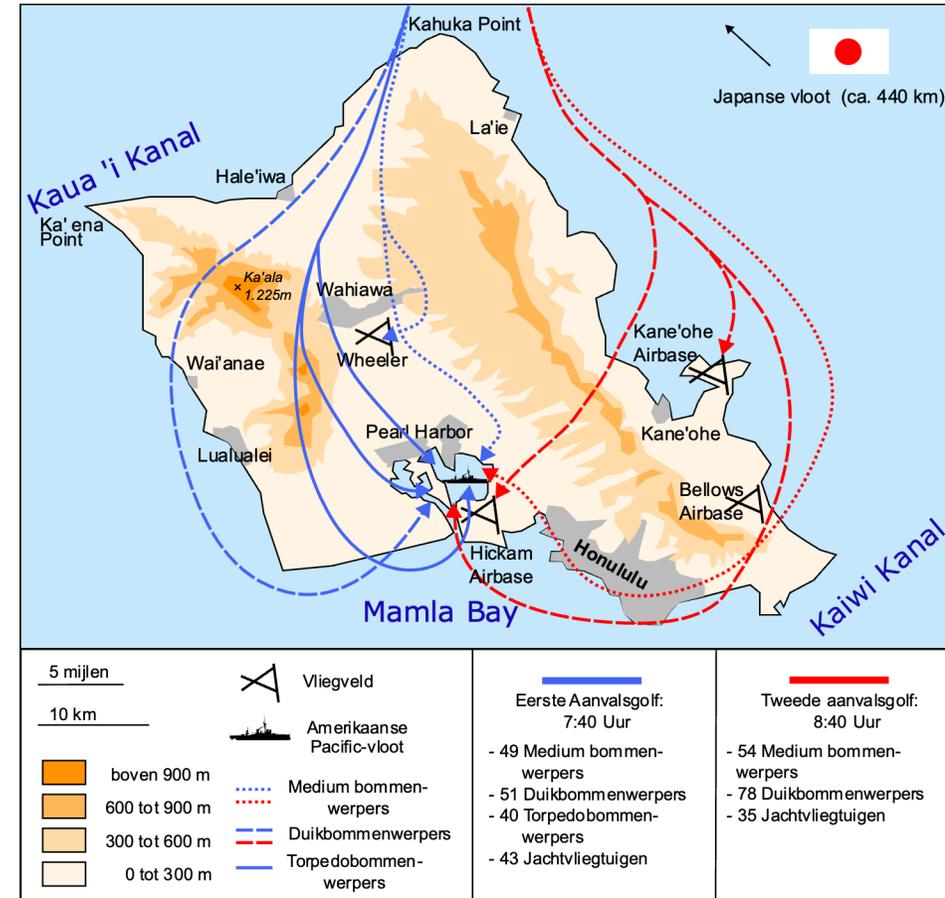


Gliederung

- Begriffseinordnung
- Wissensrepräsentation und automatisches Schlussfolgern
- Agenten und Aktionen
- **Automatisiertes Planen**

Handlungsplanung

- Einige Wörterbuch Definitionen des Begriffs „Plan“:
 1. Vorstellung von der Art und Weise, in der ein bestimmtes Ziel verfolgt, ein bestimmtes Vorhaben verwirklicht werden soll
 2. Absicht, Vorhaben



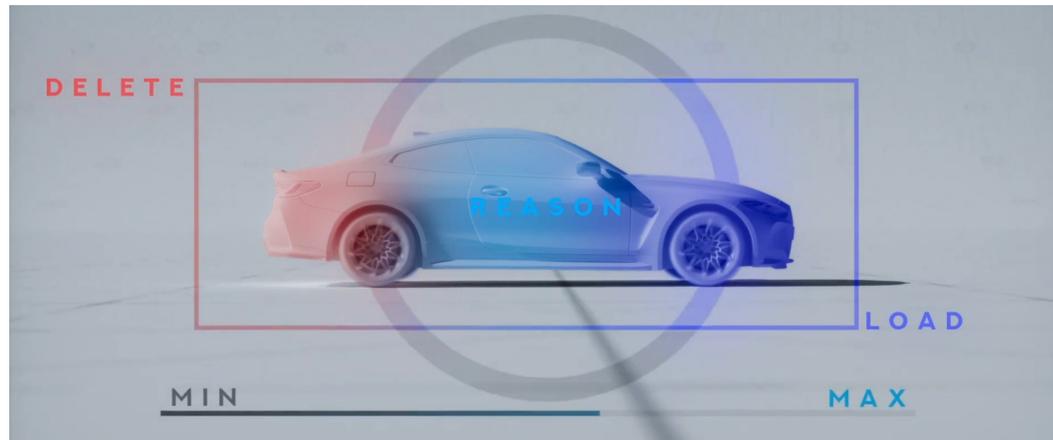
Pearl_Harbor_1941_de.svg: Thomas Steiner, User:W.wolny,
User:HistoricairPearl_Harbor_1941.svg: historicair 19:17, 9 Oct
2006 (UTC) Battleship clipart.svg: User:Citypeekderivative
work: Citypeek, CC BY-SA 3.0
via Wikimedia Commons

Handlungsplan

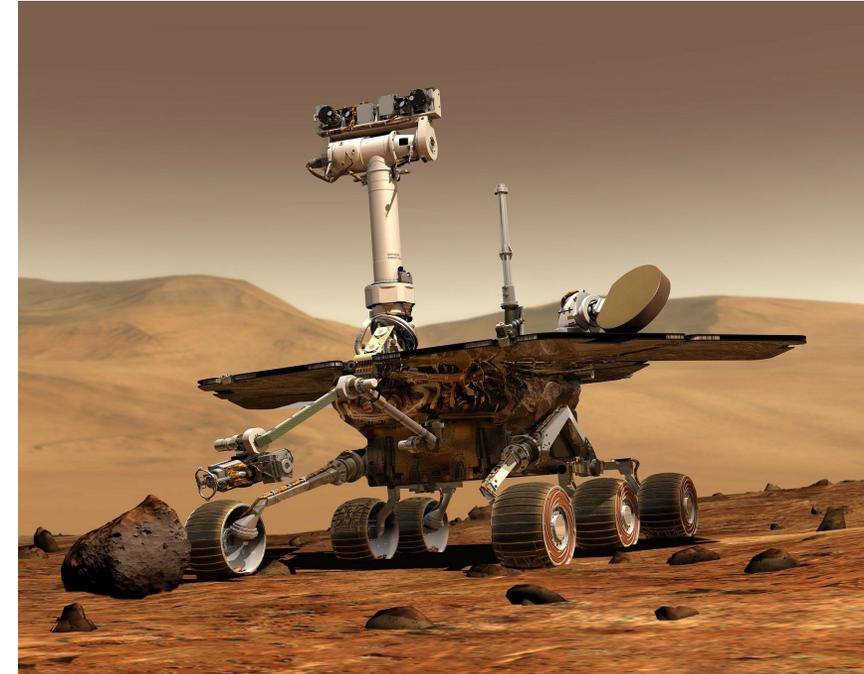
- Für uns ist ein Plan eine **Abfolge von Aktionen**, die ausgeführt werden, um (hoffentlich) ein bestimmtes **Ziel** zu erreichen
 - es ist nicht zwangsläufig so, dass die Aktionen das Ziel **wirklich erreichen werden**
- **Handlungsplanung** ist der Prozess des Findens eines Planes
 - mit hinreichend guter Qualität
- Handlungsplanung ist eng verwandt zum „Denken“
 - Menschen und einige Tiere denken nach, wenn ihnen Wahlmöglichkeiten gegeben werden

Anwendungen der Handlungsplanung

- Weltraumerforschung
 - Deep Space 1, Mars Exploration, ...
- Autonomes Fahren
- Hausautomatisierung
 - Saugroboter, Intelligente Thermostate...
- Logistik
- Lehre
- ...



Oxford Semantic Technologies



Planer

- Ein Planer ist etwas, das durch Suchverfahren Lösungen für Planungsprobleme finden kann
- In der Informatik ist das **automatische** und das **halb-automatische** (gemischt-initiative) Planen relevant
 - Planer die ohne oder mit wenig menschlicher Intervention arbeiten (z.B. nur um Präferenzen zu erfragen)

Arten von Planern

- Wir unterscheiden verschiedene Arten von (automatisierten) Planern:
 - **Domänen-spezifische**
 - können nur eine Art von Planungsproblem lösen
 - Beispiele: GPS Navigation, Schach spielen, Fertigungsroboter
 - **Domänen-unabhängige**
 - können viele verschiedene Planungsprobleme lösen
 - ähnlich zu generellen Programmiersprachen wie Java, C++, ...
 - **Konfigurierbare**
 - nutzen einen domänen-unabhängigen Planer
 - passen einige Parameter für eine bestimmte Domäne an

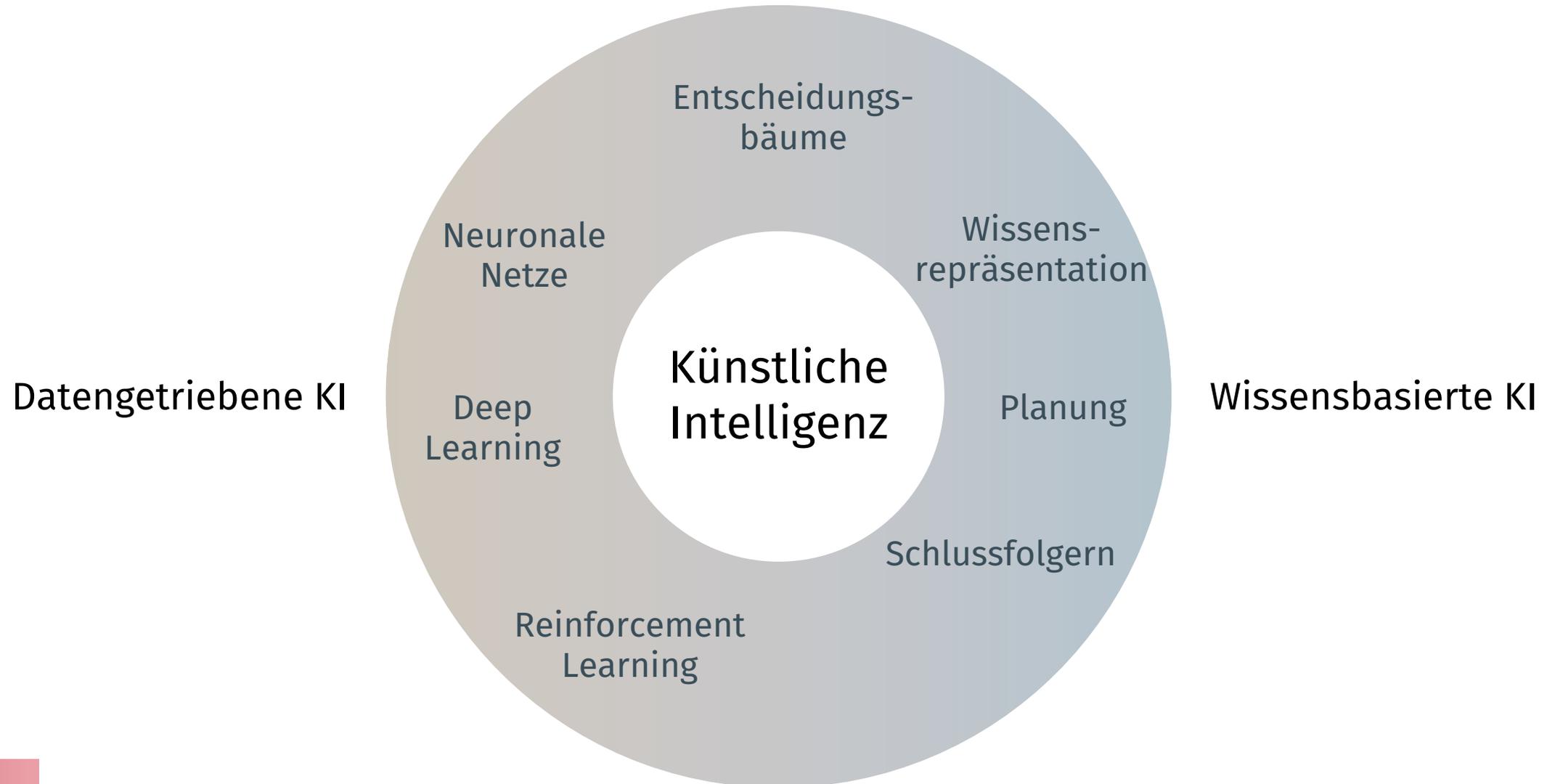
Klassisches Planen

- Viele Planungsprobleme können als **klassische, domänen-unabhängige Planungsprobleme** formalisiert werden
- Erfasst werden dazu
 - mögliche **Zustände** des Systems
 - **Aktionen**, die durchgeführt werden können
 - eine **Zustandsübergangsfunktion**: Was bewirkt eine Aktion „a“ in einem Zustand „s“?
 - der **initiale Zustand** des Systems
 - die **Zielzustände**
- (Automatisierte) Planer berechnen dann Aktionsabfolgen, um ein Ziel zu erreichen

Zusammenfassung

- In der wissensbasierten Handlungsplanung **nimmt** ein **Agent** seine **Umgebung wahr** und **führt Aktionen aus**
- Es gibt verschiedene **Typen von Agenten**
- Fortschrittliche Agenten nutzen **Zustände** um ihre Sicht auf die Welt abzubilden und ihre Aktionen sind durch **Ziele** und **Nutzenfunktionen** geleitet
- Planer bestimmen **optimale** Aktionen im **Zustandsraum**, dem Modell der Umgebung, um Ziele zu erreichen
- **Partiell beobachtbare, stochastische, dynamische** oder **kontinuierliche** Umgebungen sind deutlich schwerer zu handhaben als die hier betrachteten voll beobachtbaren, deterministischen, statischen und diskreten Umgebungen

Künstliche Intelligenz = Maschinelles Lernen?



Schlussworte

- Wir haben hier **wissensbasierte** (symbolische) Methoden betrachtet
- Vorteile:
 - Erklärbarkeit (Vertrauen), Analysierbarkeit, Vorhersagbarkeit, ...
- Argumente für lernende (subsymbolische) Methoden:
 - Robustheit gegenüber Störungen in den Daten, Performanz auf großen Daten, Behandlung von Wahrnehmungsproblemen...
- Beide Herangehensweisen haben ihre Stärken und sollten da eingesetzt werden, wo ihre Stärken liegen oder in Kombination!
- Herausforderungen sind der verantwortungsvolle Umgang mit personenbezogenen Daten, gerade bei (black-box) datengetriebenen Verfahren