

Ego cogito, ergo sum!  
Ich rechne, also bin ich?

Eine kritische Analyse von KI-Systemen aus einer  
entscheidungstheoretischen Perspektive

Sabine Theresia Köszegi



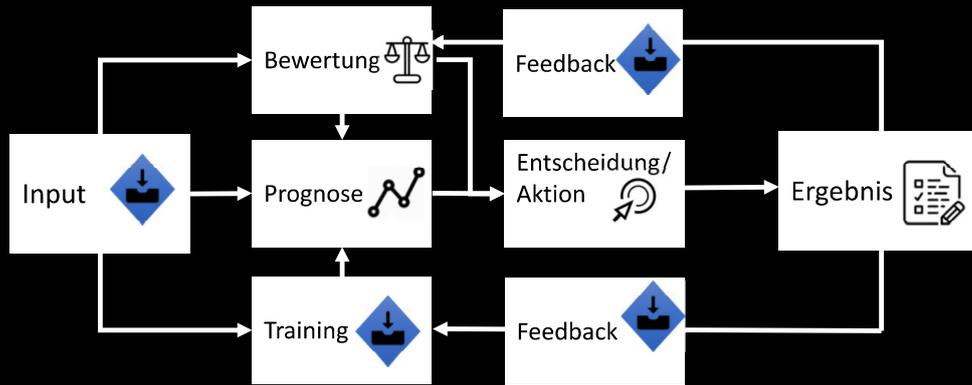
René Descartes

# Prediction Machines - Vorhersagemaschinen

- Prognose von **aktuellen oder zukünftigen Zuständen** ist zentral bei Entscheidungen unter Unsicherheit
- Technologien, die **Prognosequalität** verbessern, erhöhen **Entscheidungsqualität**
- ... und führen zu großen Hoffnungen, den Menschen bei Entscheidungen zu **unterstützung/zu ersetzen** ...
  - Bessere Diagnosen und Therapien in der Medizin
  - Effizienterer Ressourceneinsatz in Industrie und Landwirtschaft
  - Zielgerichteteres Marketing
  - Bessere Entscheidungen in der Personalauswahl ...
  - Etc.....

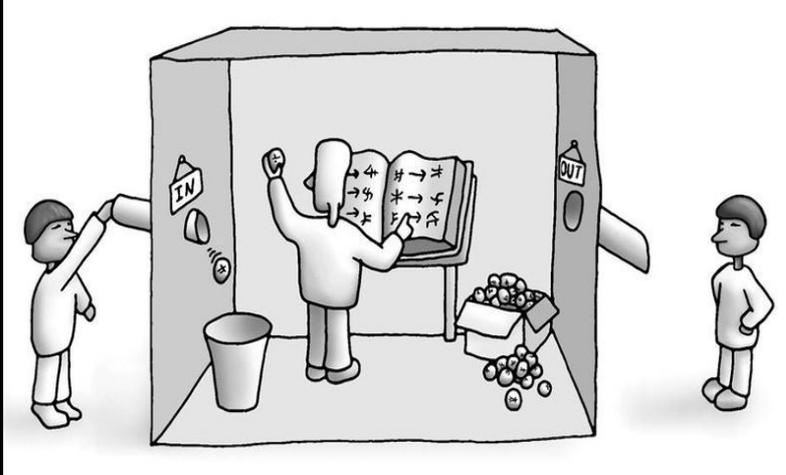
# Anatomie einer Entscheidung

◆ = DATEN



Prognose mithilfe von Maschinellem Lernen & Algorithmen (Handlungsvorschrift zum Lösen einer Aufgabe)

# Der Chinesische Raum – Searls Gedankenexperiment



Source: wikicommons

Man stelle sich einen Mann mit englischer Muttersprache vor, der kein Chinesisch spricht und in einen Raum voller Schachteln mit chinesischen Schriftzeichen (einer Datenbank) und einer Anleitung zur Handhabung der Schriftzeichen (dem Programm) eingeschlossen ist. Man stelle sich weiter vor, dass dieser Mann nun von draußen weitere chinesische Schriftzeichen erhält - nämlich Fragen auf Chinesisch (der Input), was der Mann jedoch nicht weiß. Und man stelle sich vor, dass der Mann in dem Raum nun durch Befolgen der Anweisungen im Programm jene chinesischen Schriftzeichen auswählt, mit denen die Fragen richtig beantwortet werden (der Output). Mit dem Programm kann die Person im Raum den Turing-Test in Bezug auf Verstehen des Chinesischen bestehen, in Wahrheit versteht er jedoch kein einziges Wort Chinesisch.

# Was KI-Systeme derzeit können... (RECHNEN)

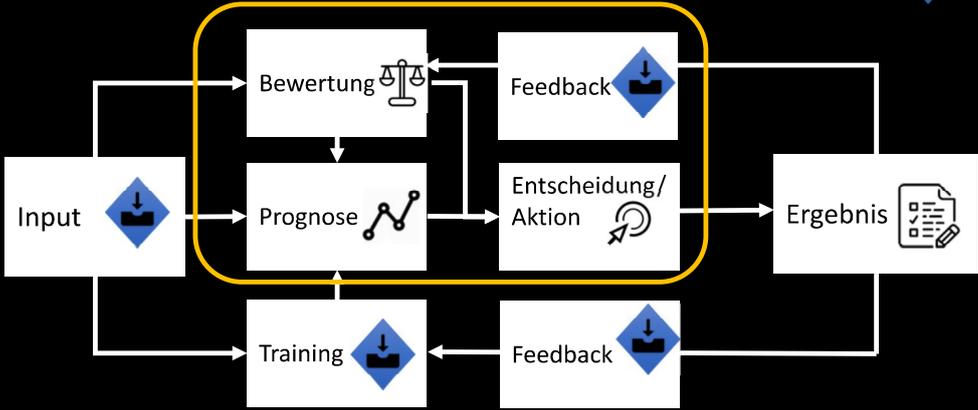
- Wahrnehmen (Sensorik/Bilder/Sprache)
- Erkennen von Mustern (Korrelationen, lernen, klassifizieren, vorhersagen)
- Anwenden von Regeln (Logik, Reasoning)
- Ausführen (auf Basis definierter Ziele, nachahmen)

# Was Menschen können... (alles was Maschinen können und)

- Bewerten (Ziele, Wünsche, Bedürfnisse, Werte, Ethik & Moral)
- Verstehen (Kontext, Sinn, Bedeutung, Humor, Satire, Zynismus)
- Fühlen (Emotionen, Empathie, Ästhetik, Genuss)
- Kreativität (Innovation, Spontaneität)
- Verlernen (loslassen, vergessen, verzeihen)

# Humanzentrierte Entscheidungen

◆ = DATEN



Adaptiert von : Agrawal, Gans & Goldfarb, 2018

Entscheidungen, die Maschinen für uns Menschen treffen,  
treffen wir nicht selbst!

Je mehr Entscheidungen wir an Maschinen delegieren, desto  
weniger Entscheidungen treffen wir selbst!

Wir sollten daher uns daher genau fragen, welchen Preis wir  
dafür bezahlen...

in Hinblick auf Autonomie & Selbstbestimmung

# Prinzipien für KI Systeme – Ethikrichtlinien



Autonomie



Fairness

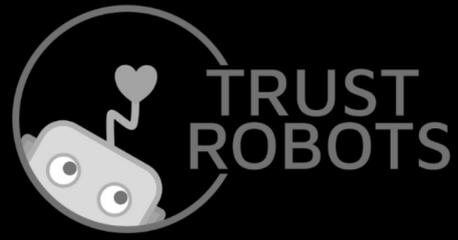


Erklärbarkeit



Sicherheit

# Fragen?



Sabine Theresia Köszegi  
[sabine.koeszegi@tuwien.ac.at](mailto:sabine.koeszegi@tuwien.ac.at)