

Ethik und KI

Dr. Aljoscha Burchardt

ASTT-TAGUNG 2022: EINE WELT OHNE BABEL?



Deutsches Forschungszentrum für
Künstliche Intelligenz GmbH

KI für den Menschen.



By way of a definition...

KI Systeme kommen einem intelligent vor.



KI Systeme haben in der Lösung der Aufgabe eine gewisse Freiheit.





Wie funktioniert KI?

Motivation

Implicit knowledge



Explicit knowledge



AI System

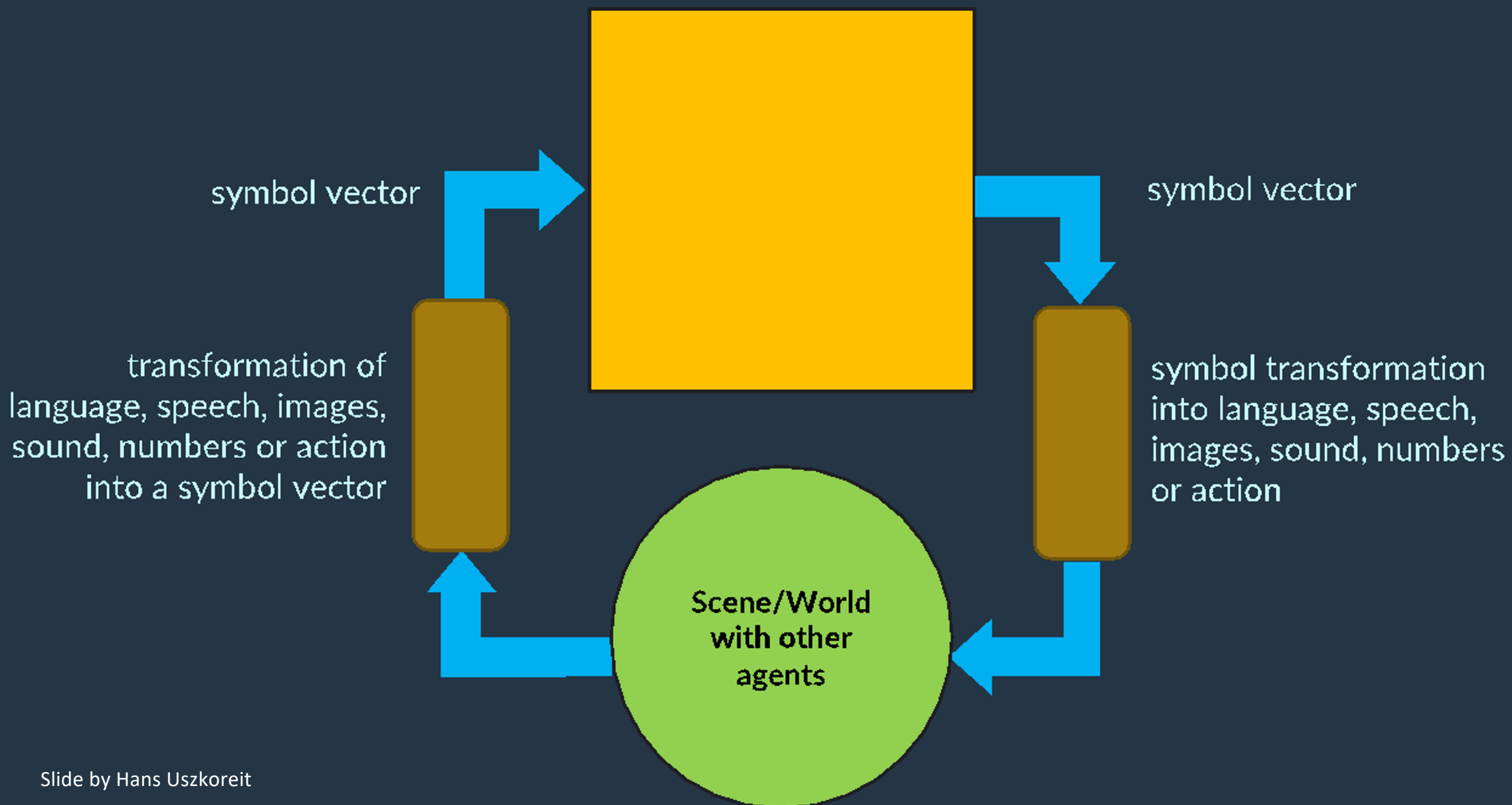
symbol vector

symbol vector

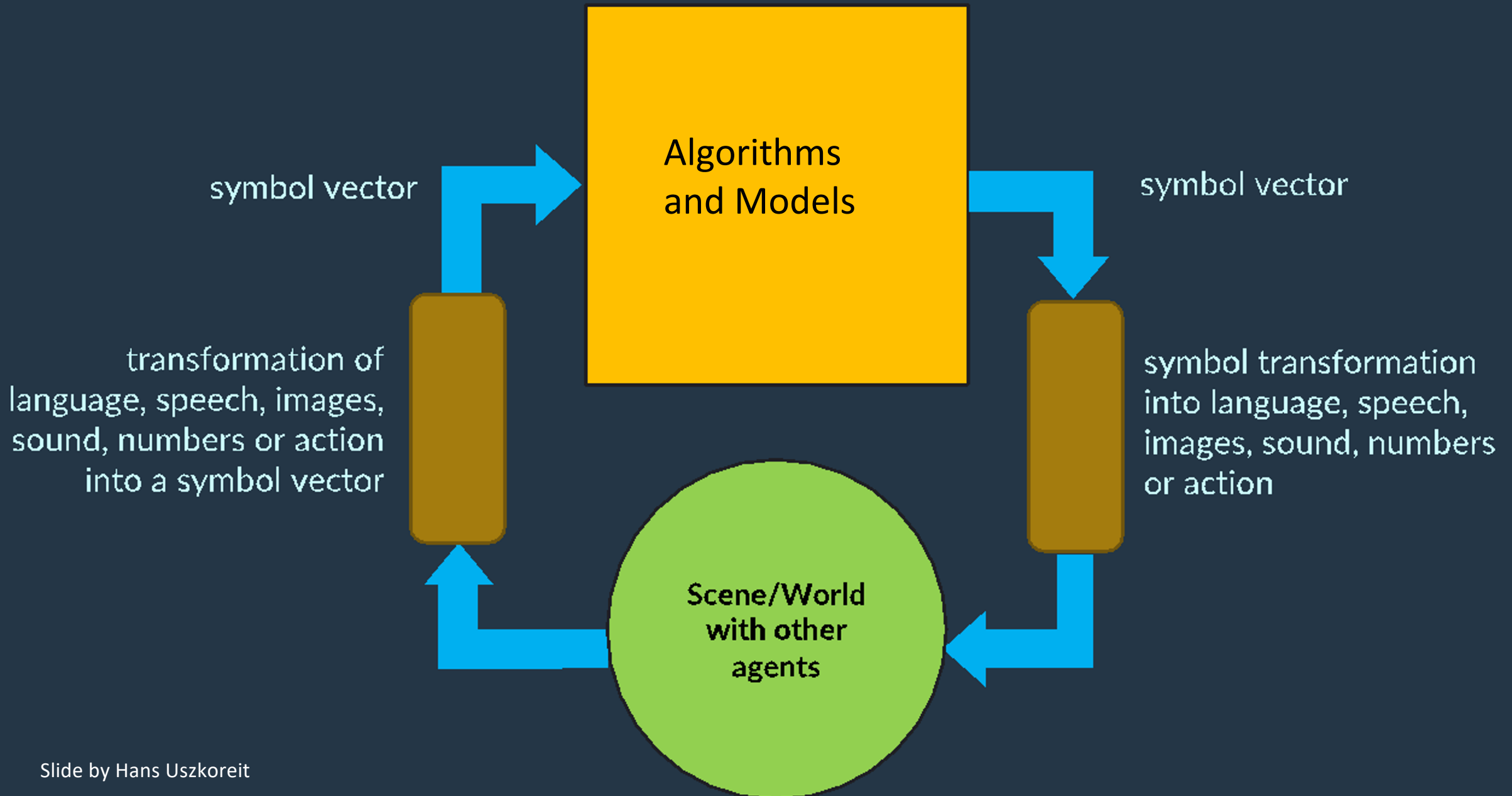
transformation of
language, speech, images,
sound, numbers or action
into a symbol vector

symbol transformation
into language, speech,
images, sound, numbers
or action

Scene/World
with other
agents



AI System



symbol vector

Algorithms and Models

symbol vector

transformation of language, speech, images, sound, numbers or action into a symbol vector

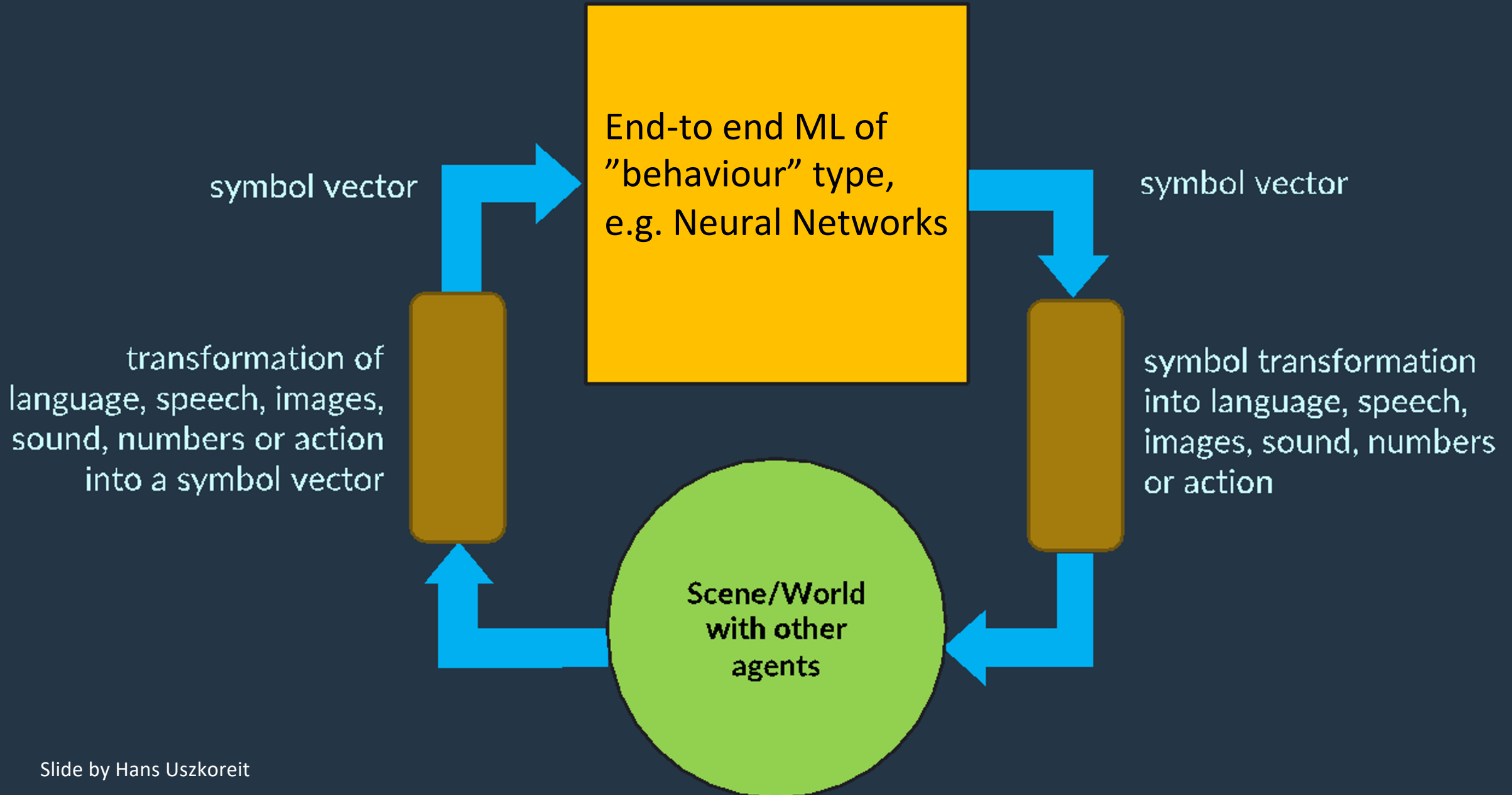
symbol transformation into language, speech, images, sound, numbers or action

Scene/World with other agents

Algorithmus vs. Black Box Models

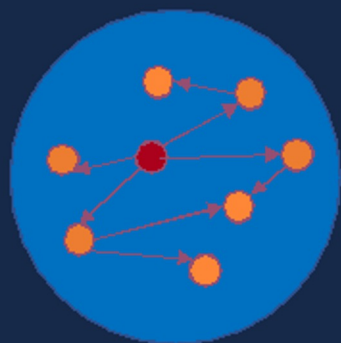


AI System



input vector may represent:

- an image
- a sentence in English
- a sentence in German
- a speech recording
- sales figures
- camera input



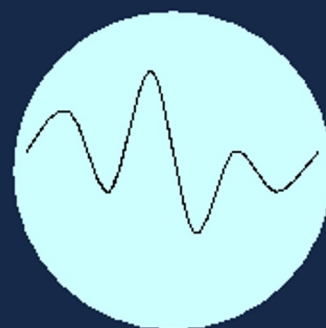
input symbol vector



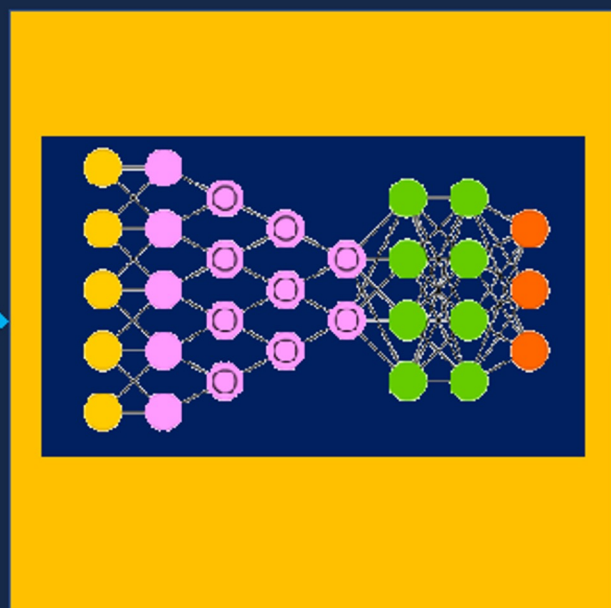
training data	labels
input vector 1	label vector 1
input vector 2	label vector 2
input vector 3	label vector 3
input vector 4	label vector 4
input vector 5	label vector 5
input vector 6	label vector 6

label vector may represent

- a type of object
- a sentence in English
- a sentence in German
- a spoken phrase
- sales figures
- driving action

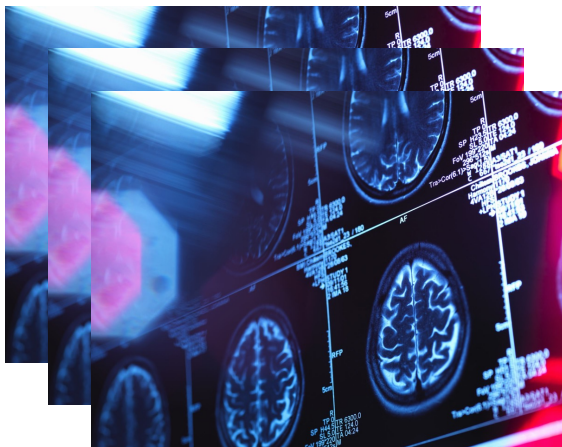


output symbol vector



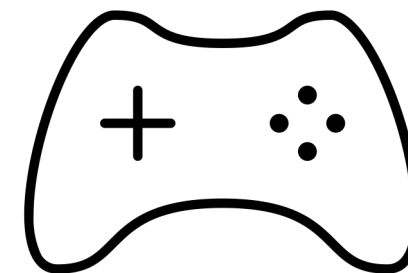
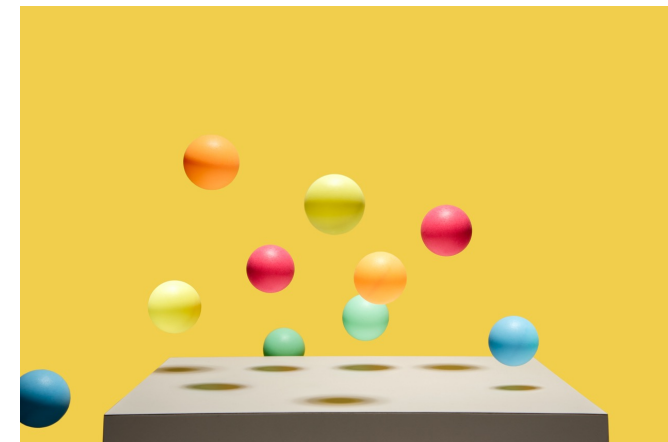
Machine Learning Examples

Supervised



Tumor: yes
 Tumor: yes
 Tumor: no

Reinforcement Learning



Score: 74

Hallo Welt.	Hello world.
This is an example.	Dies ist ein Beispiel.
And so on.	Und so weiter.

Begriffsklärung: Ethik vs. Moral

Individuelle Ebene	Reflexionsebene
Hans handelte moralisch.	Maria studiert Ethik.
Maria verhielt sich sozial.	Hans studiert Soziologie.

- Moral: normative (nicht unbedingt kodifizierten) Regeln, Sitten und Gebräuche in einer gegebenen Gemeinschaft (nicht: rechtlich; nicht: reine Konventionen wie Höflichkeit).
- Ethik: Wissenschaft von der Moral, als Reflexionswissenschaft
- (Es gibt noch andere Verwendungen der Begriffe.)

Vorsicht: false friend!



ethical \cong moralisch

Was kann „ethische KI“ bedeuten?

1. Entwicklung von Systemen, die moralisch gute Zwecke realisieren (sollen).
2. In den Entwicklungs- und Implementationsprozess derartiger Systeme oder in deren Nutzung fließen ethische Überlegungen ein.
3. KI-basierte Systeme werden zu dem Zweck entwickelt und implementiert, ethische Deliberation, Begründung und die moralische Entscheidungsfindung zu unterstützen bzw. zu erleichtern (“Moral Machine“).



Bild: <https://www.jpc.de>

Rahmen für eine vertrauenswürdige KI

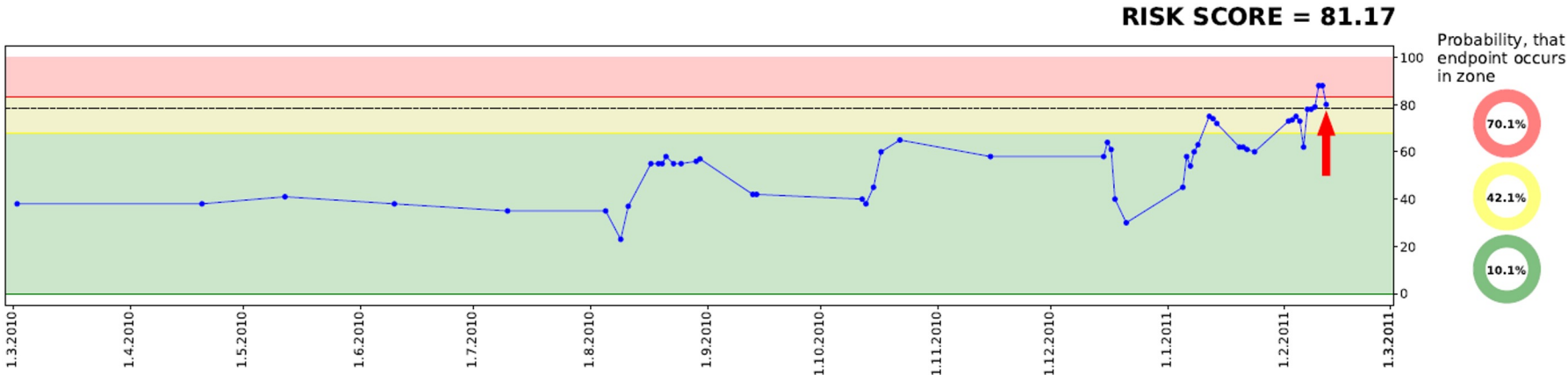
- EU HLEG on AI



Data & Model



A dashboard with risk score, traffic light system and feature relevance



Most influential features for this decision:

- Feature 1 with value (relevance: 22.49)
- Feature 2 with value (relevance: 9.09)
- Feature 3 with value (Relevanz: 7.88)
- Feature 4 with value (relevance: 4.66)
- Feature 5 with value (relevance: 4.63)

Most influential features for overall model:

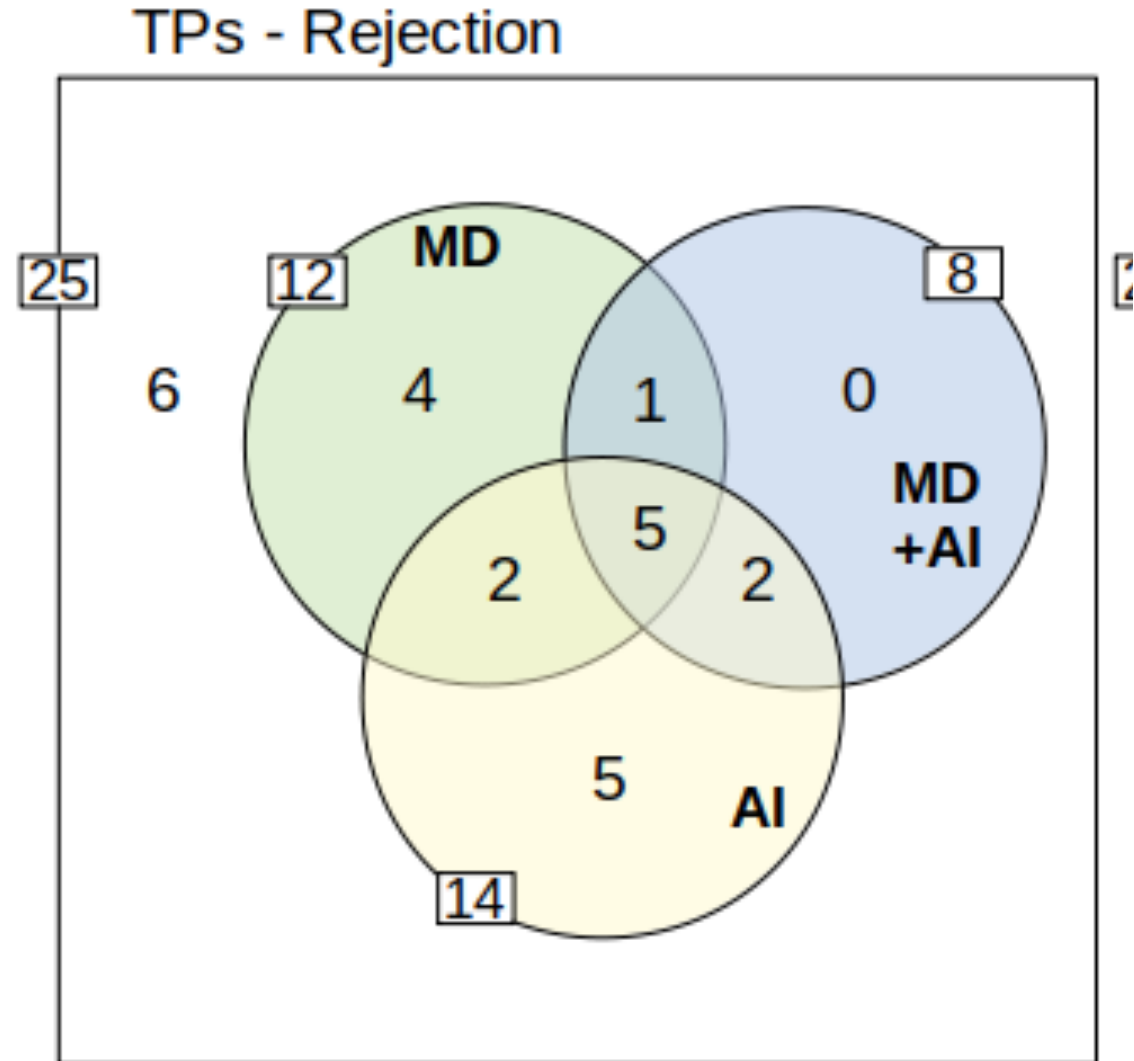
- Feature A (relevance:35.12)
- Feature B (relevance:23.76)
- Feature C (relevance:8.99)
- Feature D (relevance:3.04)
- Feature E (relevance:2.26)



AI vs MD vs MD+AI: Findings



- AI outperforms MDs
- Junior MD performs lower than Senior MD
- Junior MD increases in performance using AI, while Senior MD decreases in performance!
- MDs seem to be influenced by decision of AI
- MDs and AI partially detect different patients at risk



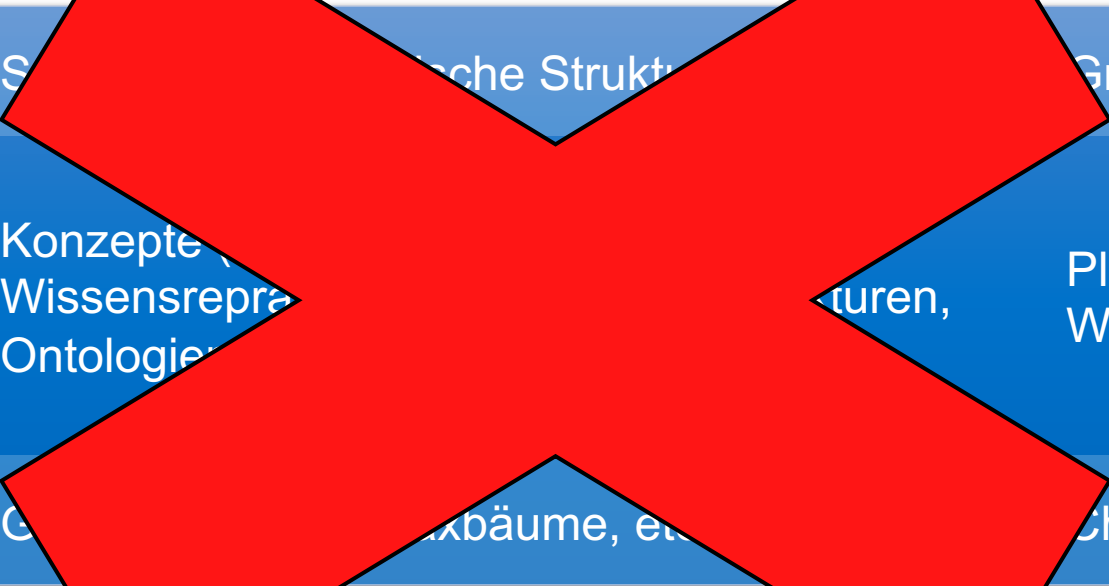
Meaningful human control



Idealistische Sicht der Computerlinguistik

Pragmatik	Sprechakte, Rhetorische Struktur, etc.	Grice, Habermas, ...
Semantik	Konzepte (Entitäten, Relationen, Events), Wissensrepräsentation (Merkmalsstrukturen, Ontologien, Logik), Inferenz, etc.	Platon, Leibnitz, ..., Wittgenstein
Syntax	Grammatiken, Syntaxbäume, etc.	Chomsky ff.

Idealistische Sicht der Computerlinguistik



Pragmatik	Sprachliche Strukturen	Grice, Habermas, ...
Semantik	Konzepte, Wissensrepräsentationen, Ontologien	Platon, Leibnitz, ..., Wittgenstein
Syntax	Grammatische Bäume, etc.	Chomsky ff.

Holzschnittartige Sicht auf NLP / Sprachtechnologie

- Flache Verarbeitung an der „Wort-Oberfläche“
 - Sehr wenig explizites Weltwissen, kaum sprachliches Wissen im Einsatz
- Trade-Off (inhaltliche) Kontrolle vs. Skalierbarkeit/Performanz
- In der Praxis im Wesentlichen drei Ansätze
 1. Symbolische Wissensverarbeitung („Datenbanken“)
 2. Klassisches Machine Learning („Bag of Words“ und Merkmale)
 3. Neuronale Netze („Word Embeddings“)

Kontrolle bei der Maschinellen Übersetzung

History of MT

- „Shifting words“
- Rule-based MT
- Translation Memories
- Classical Machine Learning
- Today's Machine Learning

History of AI

- Algorithmics
- Expert systems
- Heuristic Algorithms
- Features / Statistics
- Neural Networks

Machine Translation Example

☰ Google Übersetzer



Text



Dokumente

SPRACHE ERKENNEN

DEUTSCH

ENGLISCH ▾



ENGLISCH

DEUTSCH

FRANZÖSISCH

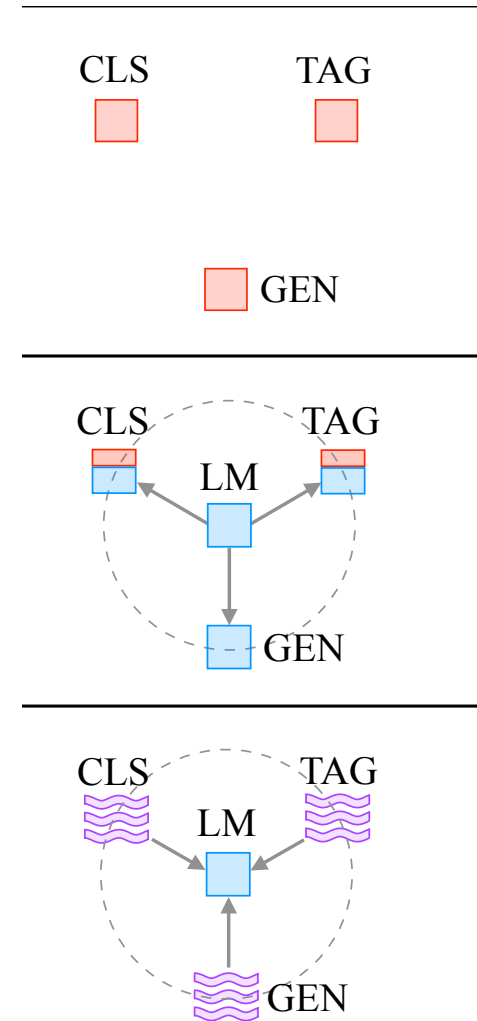
Du sollst nicht ehebrechen.
Du sollst nicht Ehe brechen.
Du sollst nicht eher brechen.
Du sollst nicht ehe brechen.



You shall not commit adultery.
You shouldn't commit adultery.
You shouldn't break sooner.
You shall not break before marriage.

Neuronale Architekturen im Wandel

- Taskspezifische Modelle, fully supervised
- Vortrainiertes Sprachmodell + Finetuning
- Vortrainiertes großes Sprachmodell + Prompting



Grafiken leicht abgewandelt aus:
<https://arxiv.org/pdf/2107.13586.pdf>

Sprachmodelle

- Aufgabe: Nächstes Wort vorhersagen
 - Trainingsdaten: Web, Mehrsprachiges, Programmcode, etc.
 - BERT
 - >100 Millionen Parameter
 - Trainingsdaten: Tausende Millionen Wörter
 - GPT-3
 - 175 Milliarden Parameter
 - Trainingsdaten: Hunderte Milliarden Wörter
 - **GPT-3** creates a carbon footprint equivalent to traveling 700,000 kilometers by car for a single training session.
- <https://www.anthropocinemagazine.org/2020/11/time-to-talk-about-carbon-footprint-artificial-intelligence/>
- According to one estimate, **training** GPT-3 would **cost** at least **\$4.6 million**.

Prompting (Luminous, Aleph Alpha)

GENERATION

Press Release

Luminous can write formatted texts from collections of information. Enter a few key points (comma-separated) in the window below and it will write a press release! But be careful: if Luminous is missing information, it may make stuff up!

Information:

May 27 2022, Aljoscha Burchardt from DFKI speaks at ESMT alumni group meeting in Berlin: "NLP is at the core of AI developments",

Press release:

May 27th, 2022
 Aljoscha Burchardt from DFKI speaks at ESMT alumni group meeting in Berlin
 Berlin. Aljoscha Burchardt from DFKI spoke at the ESMT alumni group meeting in Berlin. He presented the current state of the art of NLP and its importance for the development of artificial intelligence.

[View Settings](#)

[Open in Playground](#)

Reset

Submit

Prompting: Press Release

Write press releases with the following information separated by commas:

###

Information: November 1 2023, WWF reports that Northern white rhinoceros was cloned, Embryo transplanted into surrogate mother's (southern white rhinoceros) womb, Baby to be born in 16 months

Press release: November 1st, 2021

Northern white rhinoceros to be saved from extinction

WWF. Gland, Schweiz. WWF scientists report that for the first time, a northern white rhino has been successfully cloned and injected into a surrogate. Before, the species had become functionally extinct with the death of the last male rhinoceros. The baby will be carried out by a female of a closely related species. Presumably, it will be born in 16 months.

###

Information: October 16 2030, NASA reports touchdown on Mars, Mission "Conquer" successful, 4 astronauts on board, Team to remain for 7 days

Press release: October 16th, 2030

First manned mission landed on Mars

Washington, D.C. After decades of anticipation, the first manned mission has successfully landed on Mars. On board of the touchdown vehicle were four astronauts, who are currently investigating the red planet's surface. They are to remain on Mars for 7 days before returning into orbit. Later this year, more manned missions will head to Mars to establish a research colony.

###

Information: May 27, 2022, Aljoscha Burchardt...

Press release:

Anwendungen mit Prompting

- Frage-Antwort-System
- Leichte Sprache
- Übersetzung
- Text in Tabelle überführen
- Zusammenfassung
- (HTML-)Code erzeugen
- Einfaches Rechnen
- Einfaches Schließen
- Etc.

KI als Muse

- Frage: Was hat KI mit Zukunftsforschung zu tun?
- Antwort:
 - KI ist ein wichtiges Werkzeug, um die Zukunft der Menschheit zu erforschen.
 - Frage: Was ist die Zukunft der Menschheit?
 - Antwort: Die Zukunft der Menschheit besteht aus vielen Möglichkeiten.
 - Frage: Welche Möglichkeiten gibt es?
 - Antwort: Die Möglichkeiten sind so vielfältig, dass sie nicht alle in einem Satz erfasst werden können.
- Frage: Wie wird Künstliche Intelligenz die Zukunftsforschung verändern?
- Antwort:
 - Die Zukunft wird zunehmend durch KI-Systeme beeinflusst. Die Zukunft ist nicht mehr so, wie sie einmal war.
 - Frage: Wie sieht die Zukunft aus, wenn KI-Systeme die Zukunft beeinflussen?
 - Antwort: Die Zukunft wird mehr und mehr ein KI-System sein.
 - Frage: Was sind die Hauptvorteile von KI-Systemen?
 - Antwort: KI-Systeme haben die Fähigkeit, sich selbst zu verbessern.

Multimodales Kontextverstehen



This girl is happy because

Completion

she is wearing a bracelet that she made herself.

**KI-Modell: Luminous von Aleph Alpha
Bild: Aljoscha Burchardt
Modell: Klara Klenk**

Danke!

☰ Google Übersetzer

Text

Dokumente

SPRACHE ERKENNEN

TÜRKISCH

DEUTSCH

ENGLISCH



DEUTSCH

ENGLISCH

FRANZÖSISCH



der, die, das, wieso, weshalb, warum



the, the, the, why, why, why



36/5000

