

# Kurzeinführung KI (und Text Mining)

Gerhard Heyer

Universität Leipzig

heyer@informatik.uni-leipzig.de

UNIVERSITÄT LEIPZIG

Institut für Informatik



Automatische Sprachverarbeitung

## Künstliche Intelligenz – 1. Hype-Cycle

### **Grundlegende Idee in den 50er Jahren („*Dartmouth-Konferenz*“):**

- **Denken ist Symbolverarbeitung.**
- **Grundsätzlich sollte es also möglich sein, mit Programmen, die Symbole verarbeiten, natürliches Denken zu simulieren.**
- **(Annahme: zwischen einem Gehirn, das denkt, und einem Elektronengehirn, das Denken simuliert, besteht kein Unterschied, sog. *starke These der künstlichen Intelligenz*)**

### Künstliche Intelligenz – 2. Hype-Cycle (80er Jahre)

#### In den Folgejahren Fokussierung auf

- **symbolische Repräsentationen und Regelsysteme (*Logiken*)**
- **Programmiersprachen und Hardware für deren Verarbeitung (LISP, Prolog, Lisp-Maschinen, ...)**
- **Merkmalsbasierte Modellbildungen für Anwendungen der KI (*knowledge engineering*)**

**Erfolg der Anwendung hängt wesentlich von der Wahl passender Merkmale (*features*) ab.**

## Künstliche Intelligenz – 3. Hype-Cycle

### **Dank Fortschritten in der Hardwareentwicklung bestimmen seit den letzten 15 Jahren sub-symbolische *künstliche neuronale Netze* die Forschung**

- neuronale Netze lernen selber z.B. im Hinblick auf eine Klassifikationsaufgabe (Text, Bild, usw.) die besten Merkmale zu identifizieren und anzuwenden (*überwachtes Lernen*)
- Anwendungen mit neuronalen Netzen sind *datengetrieben* und erfordern hohen Aufwand beim *Trainieren* der Netze
- Transformer Architekturen ermöglichen die Transformierung einer Eingabekette von Daten in eine Ausgabekette unter Berücksichtigung des Kontextes einzelner Daten
- Für textbasierte Anwendungen stehen inzwischen große Mengen an Daten und sog. *Sprachmodelle* (vorberechnete Übergangswahrscheinlichkeiten) zur Verfügung

## Künstliche Intelligenz – neuronale Transformer Modelle

### Transformer Modelle stellen einen Paradigmenwechsel in der KI dar

- sie ermöglichen die parallele Verarbeitung sehr großer Datenmengen (Text, Bild, Video, Musik, Bewegung, ...) mit einer sehr großen Anzahl von Parametern (*features*)
- *encoder*-basierte Modelle (z.B. *BERT*, *RoBERTa*) sind darauf trainiert, ein Wort in einem vorgegebenen Kontext vorherzusagen, Anwendungen: Klassifikation, Erkennung von Entitäten, Sentiment-Analyse, ...
- *decoder*-basierte Modelle (z.B. *GPT-4*, *BLOOM*) sind darauf trainiert, das nächste Wort einer vorgegebenen Sequenz von Wörtern vorherzusagen, Anwendungen: Textgenerierung und -veränderung, Übersetzung, Fragebeantwortung, ...

### Künstliche Intelligenz – ChatGPT

**ChatGPT (generative pre-trained transformer) basiert auf sehr großen Daten mit vortrainiertem Transformer-Modell *plus* umfangreichen manuellen Korrekturen (unter Verwendung von *reinforcement learning*)**

- **Unter einer einfachen und intuitiven Benutzeroberfläche kann ChatGPT Fragen beantworten und Anweisungen ausführen und nutzt dabei die vordefinierten Texte und Übergangswahrscheinlichkeiten**
- **Anweisungen (*prompts*) können u.a. sein Klassifikationsaufgaben, Konsistenzprüfungen, *chain-of-thought prompting*, *generate knowledge prompting***
- ***prompting engineering* bietet Möglichkeiten für eine effiziente Nutzung von ChatGPT und anderen generativen Transformern**

Große Textdaten – es muss nicht immer ChatGPT sein

## Wortschatz-Projekt der Uni Leipzig

<https://wortschatz.uni-leipzig.de/de>

- **Daten seit 1994**
- **Online-Nachrichtenquellen**
- **Umfang *Deutsch* 2022: 31 Mio Sätze, 500 Mio token**
- **Anzeige von Daten in Jahresscheiben**
- **Darüber hinaus Angaben zu Grammatik und Semantik**
- **tagesaktuelle Anzeige statistisch auffälliger Themen in den „Wörtern des Tages“**

<https://wod.corpora.uni-leipzig.de/de/de>

- **Mehr Infos zu Text Mining**

<https://link.springer.com/book/10.1007/978-3-658-35969-0>

### Einsatzmöglichkeiten von KI für die Bürgerbeteiligung

#### **Voraussetzung: Klärung tatsächlicher Bedarfe!**

- **Nutzung analytischer KI für**
  - die Klassifikation von großen Textmengen, das Aufzeigen von Zusammenhängen, Identifikation von Akteuren und deren Netzwerken, Zusammenfassung von Beiträgen, ...
- **Nutzung generativer KI für**
  - die Unterstützung bei der Erstellung von Texten und Formularen, insb. auch durch Vereinfachungen oder Übersetzungen, ...

#### ***Aber***

- **Lösung mit neuronalen Netzen oft intransparent, abhängig von den verwendeten Trainingsdaten, ...**

## Einsatzmöglichkeiten von KI für die Bürgerbeteiligung

**Danke fürs Zuhören!**

**... und jetzt: Fragen und Diskussion**

# Einsatz von KI in der Bürgerbeteiligung

## Beteiligungsanlässe

- Vorhaben
- Medien
- Konfliktradar

## Beteiligungsscoping

- Akteursgruppen
- Angebote

## Prozessdesign

- Methoden
- Themen
- Informationen

## Texterstellung

- Zielgruppen
- Echtzeit

## Textauswertung

- Qualitativ->Quantitativ
- Events

## Dialoggestaltung

- Dialogpartner
- Konfliktampel
- Faktencheck

## Folgenabschätzung

- Simulationen
- Ressourcen

## Evaluation

- Wirksamkeit
- Prozessqualität