

Multimodale Mensch-Maschine-Interaktion

Prof. Dr. Jan-Torsten Milde
SoSe 2025

Multimodale Mensch-Maschine-Interaktion

- Dauer: 12/10 Wochen, Format: 1.5 Stunden Vorlesung + 1.5 Stunden praktische Übungen pro Woche
 - Vorlesung: Montag, 15.30-17.00 Uhr in 46.012
 - Übung: Dienstag, 15.30-17.00 Uhr in 46.012
- Ziele: Die Studierenden sollen
 - die Grundlagen und aktuellen Entwicklungen der multimodalen Mensch-Maschine-Interaktion verstehen,
 - relevante KI-Technologien kennenlernen und ein
 - eigenes prototypisches System entwickeln und implementieren, das multimodale Interaktion ermöglicht.
- Bewertungskriterien:
 - Regelmäßige Teilnahme und aktive Mitarbeit (Diskussionen und Übungen)
 - Zwischenpräsentationen/Berichte zu Experimenten mit KI-Anwendungen
 - Abschließende Präsentation und Demonstration des Prototyps
 - Dokumentation des Prototyps (Konzept, Design, Implementierung, Evaluation)

Semesterplan

Datum	Mo (WP)	Mo (MMHCI)	Fr (BegS)	Anmerkung
14.4.	+	+	-	Karfreitag
21.4.	-	-	+	Ostermontag
28.4.	+	+	-	Kommission
05.5.	+	+	+	Kommission, ab 9.50
12.5.	+	+	+	—
19.5.	+ (MR)	-	+	ACHI, Nizza
26.5.	+	+	+	—
02.6.	+	+	+	Präsentation FE bis 8.6.
09.6.	-	-	+	Pfingstmontag, Erfindermesse
16.6.	+	+	+	—
23.6.	+	+	+	Live-Coding
30.6.	+	+	+	—
07.7.	+ (MR)	-	-	SMC, Graz
14.7.	+	+	+	—

Themenüberblick

- Woche 1: Einführung und Grundlagen
- Woche 1/2: Sprachverarbeitung (Natural Language Processing – NLP) und LLMs
- Woche 3: Computer Vision
- Woche 4: Sprachsynthese (Text-to-Speech - TTS) und Auditive Interaktion
- Woche 5: Gestenerkennung und Haptische Interaktion
- Woche 6: Multimodale Fusion und Fission
- Woche 7: Kontextbewusstsein und Personalisierung
- Woche 8: Architekturen multimodaler Systeme und User Interface Design
- Woche 9: Aktuelle KI-Anwendungen in der MMI (Teil 1)
- Woche 10: Aktuelle KI-Anwendungen in der MMI (Teil 2)
- Woche 11: Prototypenentwicklung (Fokus Implementierung)
- Woche 12: Prototypenentwicklung und Abschluss

Woche 1

Einführung und Grundlagen

- Multimodale Mensch-Maschine-Interaktion
 - Definition nach Gemini: „Unter multimodaler Mensch-Maschine-Interaktion (MMI) versteht man die Interaktion zwischen Mensch und Computer, bei der mehrere verschiedene Eingabe- und/oder Ausgabemodalitäten gleichzeitig genutzt werden, um die Kommunikation natürlicher, effizienter und intuitiver zu gestalten“
 - Der Mensch verwendet eine Vielzahl von Modalitäten zur **Wahrnehmung** und **Kommunikation**.
 - Im Kontext der **multimodalen** Mensch-Maschine-Interaktion liegt der Fokus oft auf Modalitäten, die sich gut für die **Interaktion mit technischen Systemen eignen**, wie z.B. Sprache, Sehen (Gesten- und Objekterkennung), Tasten (Touchscreens, haptisches Feedback) und in Zukunft möglicherweise verstärkt auch andere sensorische Eingaben.
- Was ist KI ?
 - Siehe Foliensatz „Schüler Tag“
- Sprache und Computer
 - Das heutige Thema

Modalitäten des Menschen

- Wahrnehmung (Sinne):

- **Visuell:** Sehen (über die Augen)
- **Auditiv:** Hören (über die Ohren)
- **Taktil/Haptisch:** Tasten, Berührung, Druck, Vibration (über die Haut)
- **Olfaktorisch:** Riechen (über die Nase)
- **Gustatorisch:** Schmecken (über die Zunge)
- **Propriozeptiv:** Körpergefühl, Wahrnehmung der eigenen Körperhaltung und Bewegung (über Rezeptoren in Muskeln, Gelenken und Sehnen)
- **Vestibulär:** Gleichgewichtssinn (im Innenohr)
- **Thermozeption:** Temperaturwahrnehmung (über die Haut)
- **Nozizeption:** Schmerzempfindung (über Nervenendigungen im ganzen Körper)

- Kommunikation:

- **Sprache** (verbal): Gesprochene Worte, Tonfall, Sprachmelodie
- **Schrift** (textuell): Geschriebene Buchstaben, Symbole, Zeichen
- **Gestik:** Handbewegungen, Armbewegungen
- **Mimik:** Gesichtsausdrücke
- **Körpersprache:** Körperhaltung, Blickkontakt, räumliches Verhalten
- **Prosodie** (paraverbal): Sprechtempo, Lautstärke, Pausen
- **Haptische Kommunikation:** Berührungen zur nonverbalen Übermittlung von Botschaften
- **Piktogramme und Symbole:** Visuelle Zeichen zur Informationsübertragung

Praktische Übung: Technik/Wordle

- Themen
 - Diskussion über Beispiele für multimodale Interaktionen im Alltag.
 - Kennenlernen der Kursumgebung und relevanter Software-Tools
 - python mit virtual environment
 - pip
 - bash
 - git
 - Wortverarbeitung mit regulären Ausdrücken

Technik

- Wir wollen Bilder generieren
 - Dazu verwenden wir Fooocus
 - <https://github.com/lllyasviel/Fooocus.git>
- Aufgabe:
 - Installieren Sie Fooocus und generieren Sie ein Bild von einem Avatar mit pinken Haaren.

Sprachverarbeitung

Computerlinguistische Grundlagen

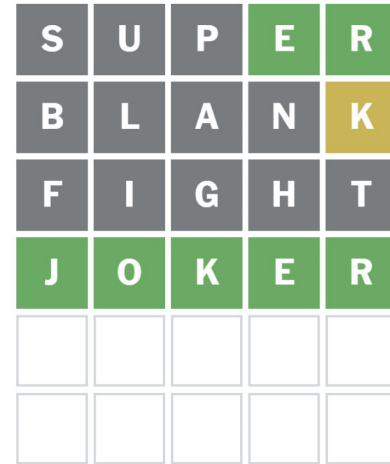
Übung: Wortanalyse

- Agenda
 - Das Wordle Problem
 - Wortbildung im Deutschen
 - Erzeugung eines Lexikon
 - Reguläre Ausdrücke
 - Größenabschätzung
 - Häufigkeitsverteilungen
 - Endliche Automaten
 - Klassifikator auf Basis der Daten

Wordle

- Einfaches Spiel, bei dem ein (englisches) Wort mit 5 Buchstaben erraten werden muss
 - Farbkodierung markiert Buchstaben

Wordle



Erste Fragen

- Wieviele deutsche Worte mit 5 Buchstaben existieren ?
- Wie kann man ein korrektes deutsches Wort erkennen und somit von einem „fehlerhaften“ Wort unterscheiden ?
- Wie kann man (schnell) ein Lexikon mit deutschen Worten erstellen ?

Kombinationen

- Um die Anzahl der Worte abzuschätzen betrachten wir die Gesamtzahl aller Buchstabenkombinationen für ein Wort der Länge 5
 - Beobachtung: es existieren $26 + 4$ Buchstaben im Deutschen (Groß-Kleinschreibung werden ignoriert)
 - Das Wort hat eine Länge von 5
 - Zeicheninventar (30) und Wortlänge (5) sind endlich
 - Hieraus folgt: es kann auch nur endlich viele Buchstabenkombinationen geben

Kombinationen

- Das Wort XXXX der Länge 5 hat dann
 - $X = 30 * X = 30 * X = 30 * X = 30 * X = 30$
mögliche Kombinationen
 - Also 30^5 Kombinationen
 - Das schätzen wir ab mit
 - $30^5 < 32^5$
 - $32^5 = 2^{5*5} = 2^{25} = 2^{10} * 2^{10} * 2^5$
 - $= (1024 * 1024) * 32$
 - Also ungefähr 32.000.000 (32 Millionen) Buchstabenkombinationen
 - Aber: davon sind die allermeisten Kombinationen kein deutsches Wort
 - Fragt sich nur: welche davon?

Übung: Lexikon

- Aufgabe
 - Erstellen Sie ein Lexikon mit deutschen Wörtern mit 5 Buchstaben
 - Nutzen Sie dazu VSCode und reguläre Ausdrücke
 - Wo bekommen Sie die Daten her ?
 - Welche Vorverarbeitung ist notwendig ?
 - Wie groß muss ein Lexikon sein ?
 - Was, außer der Wortform, könnte noch im Lexikon stehen ?
 - Arbeiten mit dem Lexikon
 - Woran erkenne ich ein Wort des Deutschen ?
 - Wie kann ich Eigenschaften von deutschen Wörtern algorithmisch/regelbasiert erkennen ?
 - Wodurch unterscheidet sich ein Wort einer anderen Sprache vom Deutschen ?

Woche 2

Sprachverarbeitung (Natural Language Processing – NLP) und LLMs

Computer Vision

Sprachsynthese (Text-to-Speech - TTS) und Auditive Interaktion

Gestenerkennung und Haptische Interaktion

Multimodale Fusion und Fission

Kontextbewusstsein und Personalisierung

Architekturen multimodaler Systeme und User Interface Design

Aktuelle KI-Anwendungen in der MMI (Teil 1)

Aktuelle KI-Anwendungen in der MMI (Teil 2)

Prototypenentwicklung (Fokus Implementierung)

Prototypenentwicklung und Abschluss