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

https://www.nytimes.com/games/wordle/index.html

Erste Fragen

- Wieviele deutsche Worte mit 5 Buchstaben existieren ?
- Wie kann man ein korrektes deutsches Wort erkennen und somit von einem „fehlerhaften“ Wort untercheiden ?
- 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

- Ausfgabe
  - 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 ?

Weitere Wochen

Sprachverarbeitung (Natural Language Processing – NLP) und LLMs

- Vorlesung:
  - Grundlagen der Sprachverarbeitung: Tokenisierung, Parsing, semantische Analyse.
  - Überblick über aktuelle NLP-Modelle: Bag-of-Words, TF-IDF, Word Embeddings (Word2Vec, GloVe, FastText).
  - Einführung in Transformer-basierte Modelle (z.B. BERT, GPT).
  - Anwendungsbeispiele von NLP in der MMI (Sprachsteuerung, Chatbots, Sprachsuche).
- Praktische Übung:
  - Experimentieren mit einer NLP-Bibliothek (z.B. NLTK, spaCy) für grundlegende Textverarbeitungsaufgaben.
  - Ausprobieren eines einfachen vortrainierten Sprachmodells (z.B. über eine API).

Computer Vision

- Vorlesung:
  - Grundlagen der Bildverarbeitung: Bildmerkmale, Filter, Objekterkennung.
  - Überblick über aktuelle Computer Vision Modelle: Convolutional Neural Networks (CNNs), Region-based CNNs (R-CNNs), YOLO, Transformers für Vision.
  - Anwendungsbeispiele von Computer Vision in der MMI (Gesichtserkennung, Gestenerkennung, Objekterkennung in interaktiven Umgebungen).
- Praktische Übung:
  - Experimentieren mit einer Computer Vision Bibliothek (z.B. OpenCV, TensorFlow, PyTorch) für einfache Bildverarbeitungsaufgaben.
  - Ausprobieren eines vortrainierten Bilderkennungsmodells (z.B. über eine API).

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

- Vorlesung:
  - Grundlagen der Sprachsynthese: Konkatenative Synthese, Parametrische Synthese (z.B. WaveNet, Tacotron).
  - Akustische Merkmale und deren Verarbeitung in der MMI (z.B. Geräuscherkennung, Sprechererkennung).
  - Design von auditiven Interfaces und Sounddesign-Prinzipien.
  - Integration von Sprache und Audio in multimodalen Systemen.
- Praktische Übung:
  - Verwendung von TTS-Engines (online oder lokal).
  - Experimentieren mit der Erzeugung und Manipulation von einfachen Sounds.

Gestenerkennung und Haptische Interaktion

- Vorlesung:
  - Verschiedene Ansätze zur Gestenerkennung (bildbasiert, sensorbasiert).
  - Grundlagen der Haptik: Taktile und kinästhetische Rückmeldung.
  - Haptische Geräte und deren Integration in MMI-Systeme.
  - Anwendungsbeispiele von Gesten und Haptik in der Interaktion (z.B. virtuelle Realität, Robotik).
- Praktische Übung:
  - Experimentieren mit einer einfachen Gestenerkennungsbibliothek oder einem Sensor (falls verfügbar).
  - Untersuchung von haptischen Feedback-Mechanismen (z.B. Vibration).

Multimodale Fusion und Fission

- Vorlesung:
  - Konzepte der multimodalen Fusion: Feature-Level Fusion, Decision-Level Fusion.
  - Strategien zur Kombination von Informationen aus verschiedenen Modalitäten.
  - Konzepte der multimodalen Fission: Aufteilung von Informationen auf verschiedene Ausgabemodalitäten.
  - Design von kohärenten multimodalen Interaktionen.
- Praktische Übung:
  - Entwurf eines einfachen Fusions- oder Fissionsmechanismus für ein hypothetisches Szenario.
  - Diskussion über die Vor- und Nachteile verschiedener Fusionsstrategien.

Kontextbewusstsein und Personalisierung

- Vorlesung:
  - Bedeutung von Kontext in der MMI (Benutzerkontext, Umgebungs Kontext, Aufgabenkontext).
  - Methoden zur Kontextmodellierung und -erfassung (z.B. Sensoren, Benutzerprofile).
  - Personalisierungstechniken in multimodalen Systemen (z.B. adaptive Interfaces, Empfehlungssysteme).
  - KI-basierte Ansätze für kontextbewusste und personalisierte Interaktion.
- Praktische Übung:
  - Brainstorming von Kontextfaktoren für das eigene Projekt.
  - Entwurf eines einfachen Mechanismus zur Berücksichtigung eines Kontextfaktors.

Architekturen multimodaler Systeme und User Interface Design

- Vorlesung:
  - Übersicht über typische Architekturen für multimodale Systeme.
  - Komponenten und Interaktionen in multimodalen Systemen.
  - Spezifische Herausforderungen im User Interface Design für multimodale Interaktionen.
  - Gestaltungsrichtlinien und Best Practices für multimodale Interfaces.
- Praktische Übung:
  - Erste Skizzen und Wireframes für das User Interface des Abschlussprojekts unter Berücksichtigung der Multimodalität.
  - Diskussion von Designentscheidungen im Hinblick auf Usability und Effektivität.

Aktuelle KI-Anwendungen in der MMI (Teil 1)

- Vorlesung:
  - Detaillierte Vorstellung aktueller KI-Anwendungen in verschiedenen Bereichen der MMI:
  - Intelligente Assistenten (z.B. Alexa, Google Assistant, Siri) und ihre multimodalen Fähigkeiten.
  - Multimodale Chatbots und Dialogsysteme.
  - KI für Gesten- und Bewegungsverfolgung in interaktiven Umgebungen.
  - Analyse der zugrundeliegenden KI-Modelle und Architekturen.
- Praktische Übung:
  - Experimentieren mit den multimodalen Fähigkeiten eines bestehenden intelligenten Assistenten oder einer anderen KI-Anwendung.
  - Analyse der Stärken und Schwächen der Interaktion.

Aktuelle KI-Anwendungen in der MMI (Teil 2)

- Vorlesung: Fortsetzung der Vorstellung aktueller KI-Anwendungen:
  - Multimodale Interaktion in Virtual und Augmented Reality.
  - KI-gestützte multimodale Interaktion in Robotik.
  - Anwendungen im Bereich Healthcare und Assistive Technologies.
  - Kreative Anwendungen multimodaler KI (z.B. in der Kunst).
  - Diskussion ethischer und gesellschaftlicher Auswirkungen dieser Technologien.
- Praktische Übung:
  - Recherche und Präsentation von weiteren aktuellen KI-Anwendungen im Bereich MMI durch die Studierenden (kurze Impulsvorträge).

Prototypenentwicklung (Fokus Implementierung)

- Vorlesung:
  - Best Practices für die Softwareentwicklung von MMI-Systemen.
  - Überblick über relevante Frameworks und Bibliotheken für die Prototypenentwicklung.
  - Tipps und Tricks für die Integration verschiedener Modalitäten in einem System.
  - Methoden zur (einfachen) Evaluation von Prototypen.
- Praktische Übung:
  - Beginn der Implementierung der Prototypen in Kleingruppen oder einzeln.
  - Erste Integration von mindestens zwei Modalitäten.

Prototypenentwicklung und Abschluss

- Vorlesung:
  - Abschließende Hinweise zur Prototypenentwicklung und -dokumentation.
  - Tipps für die Präsentation der Projekte.
  - Diskussion über zukünftige Trends und Herausforderungen in der MMI.
- Praktische Übung:
  - Fertigstellung der Prototypen (Ende des Semesters)
  - Vorbereitung der Abschlusspräsentationen (Ende des Semesters)
  - Erste informelle Präsentationen und Feedback-Runden.