

Hauptseminar

Thema: Speech To Text Verarbeitung auf lokaler Hardware

Beschreibung:

Für die Realisierung von interaktiven Assistenzrobotern wird eine leistungsfähige Sprachverarbeitungs pipeline benötigt, welche möglichst lokal auf dem Roboter PC läuft.

In anderen Hauptseminararbeiten wurden/werden dafür bereits die Sprachausgaben (Text to Speech) und die Verarbeitung der erkannten Texte mit LLMs untersucht.

Im Rahmen dieses HS soll die Spracheingabe, also ein Speech to Text System, recherchiert und auf dem Roboter evaluiert werden.

Dabei sollen aktuelle Deep Learning basierte open source Lösungen bevorzugt werden.

Als Ansatzpunkt für die Recherche kann das ASR Modell Whisper und darauf aufsetzende Projekte [1] dienen.

Aufgabenstellung:

- Recherche zu lokalen Speech to Text Systemen
- Systematisierung der verschiedenen Ansätze
- Darstellung und Beschreibung eines ausgewählten Verfahrens
- Erprobung der Software auf Roboterhardware in realistischen Einsatzbedingungen
- Präsentation der Ergebnisse im Rahmen der Hauptseminarsvorträge

Geeignet für:

- Bachelor-/ Masterstudiengänge

Themengebiet / Schwerpunkte:

- Deep-Learning
- Speech to Text

Hilfreiche Voraussetzungen:

- Besuch der Vorlesung „Mensch-Maschine-Interaktion“ und „Deep-Learning“

Literatur:

[1] <https://github.com/ochen1/insanely-fast-whisper-cli>

Betreuer: Dr. Steffen Müller (steffen.mueller@tu-ilmenau.de)
Betr. Hochschullehrer: Prof. Dr. H.M. Groß
Bearbeiter: offen